

ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

Dipartimento di Informatica – Scienza e Ingegneria (DISI)

DOTTORATO DI RICERCA IN
COMPUTER SCIENCE AND ENGINEERING
CICLO XXXII

Settore Concorsuale: 09/H1
Settore Scientifico Disciplinare: ING-INF/05

**Towards a Boolean network-based
Computational Model
for Cell Differentiation
and its applications to Robotics**

Presentata da
MICHELE BRACCINI

Coordinatore Dottorato
Prof. DAVIDE SANGIORGI

Supervisore
Prof. ANDREA ROLI

Tutor
Prof. MIRKO VIROLI

ESAME FINALE ANNO 2020

To Tamara, Amelia
and my whole
family.

Acknowledgements

Firstly I wish to thank my supervisor, Andrea Roli, for the continuous support during my PhD, for his patience, motivation, and broad and deep knowledge. His guidance illuminated my research path. I also thank Roberto Serra, Marco Villani and Sara Montagna for having collaborated with me and for sharing with me their enlightening ideas and insightful observations.

I am grateful to Sui Huang for giving me the great opportunity of working under his supervision at the Institute for Systems Biology (ISB). I am also grateful to Stuart Kauffman and Timoteo Carletti, reviewers of this dissertation, whose comments helped me improve my work.

I want to sincerely say thank to Angelo Croatti and Antonio Magnani for their friendship through these years.

I would like to thank my family, to whom I owe all I am and to whom I will be eternally grateful. I want to reserve a special mention for my girlfriend Tamara, who has constantly supported me and is capable of transforming every moment of anxiety and stress into a smile. Last but not least, I thank Amelia, my daughter on the way, my new reason for living.

Abstract

Living organisms are the ultimate product of a series of complex processes that take place within—and among—biological cells. Most of these processes, such as cell differentiation, are currently poorly understood. Cell differentiation is the process by which cells progressively specialise. Being a fundamental process within cells, its dysregulations have dramatic implications in biological organisms ranging from developmental issues to cancer formation.

The thesis objective is to contribute to the progress in the understanding of cell differentiation and explore the applications of its properties for designing artificial systems. The proposed approach, which relies on Boolean networks based modelling and on the theory of dynamical systems, aims at investigating the general mechanisms underlying cell differentiation. The results obtained contribute to taking a further step towards the formulation of a general theoretical framework—so far missing—for cellular differentiation.

We conducted an in-depth analysis of the impact of self-loops in random Boolean networks ensembles. We proposed a new model of differentiation driven by a simplified bio-inspired methylation mechanism in Boolean models of genetic regulatory networks. On the artificial side, by introducing the conceptual metaphor of the “attractor landscape” and related proofs of concept that support its potential, we paved the way for a new research direction in robotics called *behavioural differentiation robotics*: a branch of robotics dealing with the designing of robots capable of expressing different behaviours in a way similar to that of biological cells that undergo differentiation.

The implications of the results achieved may have beneficial effects on medical research. Indeed, the proposed approach can foster new questions, experiments and in turn, models that hopefully in the next future will take us to cure differentiation-related diseases such as cancer. Our work may also contribute to address questions concerning the evolution of complex behaviours and to help design robust and adaptive robots.

Keywords: *Cell Differentiation, Boolean networks, Behavioural Differentiation Robotics*

Contents

Introduction	1
I Background and Motivations	9
1 Biological Cell	11
1.1 DNA	11
1.1.1 From DNA to Protein	12
1.2 Control of Gene Expression	13
1.2.1 Regulation of Gene Expression	13
1.2.2 Gene Regulatory Networks	14
1.2.3 Epigenetic mechanisms	15
1.3 Cell Differentiation	15
1.3.1 Role of epigenetics in differentiation process	18
2 Modelling Approaches to Cellular Processes	19
2.1 What is a model?	19
2.2 Molecular biology approach	21
2.2.1 The end of central dogma?	22
2.3 Systems Biology approach	22
2.4 Complex Systems Science approach	24
2.5 Computational approaches	26
3 Cell Differentiation: challenges and complexities	29
3.1 Open questions and perspectives	29
II Dynamical systems view of Cell	33
4 Dynamical Systems	35
4.1 Attractor concept	36
4.2 Stability	37
4.3 Dynamical criticality	37
4.4 Robustness	39

5	Boolean Networks	41
5.1	Random Boolean Networks (RBN)	43
5.1.1	Dynamical regimes and criticality in RBN	43
5.1.2	Ensemble Approach	44
6	Dynamical Models of Cell Differentiation	47
6.1	Cells as dynamical systems	47
6.2	Waddington's Epigenetic Landscape	48
6.3	Cell fates as attractors	50
6.4	Cancer attractors	51
6.5	A BN-based model for cell differentiation	51
6.5.1	Attractor transition matrix (ATM)	52
6.5.2	Threshold Ergodic Set (TES)	53
III	Extensions to Current Models	55
7	Extensions to BN-based differentiation models	57
7.1	Automatic Design of Boolean Networks for Cell Differentiation	57
7.1.1	Adaptive walk algorithm	59
7.1.2	VNS-like algorithm	60
7.1.3	Results	62
7.1.4	Conclusion and future work	63
7.2	A comparison between TESs and stochastic simulation of Boolean networks	66
7.2.1	Experimental setting	67
7.2.2	Results	68
7.2.3	Conclusion	71
7.3	A software library: <i>differeNTES</i>	76
8	A network motif in BN: Self-loops	79
8.1	The impact of self-loops in RBN dynamics: A simulation analysis	79
8.1.1	Motivation and goal	79
8.1.2	Methods	81
8.1.3	Results	83
8.1.4	RBNs with OR functions	84
8.1.5	Discussion and Conclusion	86
8.2	Self-loops favour diversification and asymmetric transitions between attractors	88
8.2.1	Motivation and goal	89
8.2.2	Methods	89
8.2.3	Diversification and Asymmetry	90
8.2.4	Self-loops	91

8.2.5	Results	92
8.3	Conclusion	97
8.4	The impact of self-loops on BN attractor landscape and im- plications for cell differentiation modelling	98
8.4.1	Motivation and Goal	99
8.4.2	Methods	101
8.4.3	Results	104
8.4.4	Average number of attractors as a function of self-loops	105
8.4.5	Distribution of attractor number	108
8.4.6	Attractor robustness	111
8.5	Discussion and Conclusion	113
9	Wrap-up	117
10	Epigenetics-driven differentiation in BN models	119
10.1	State of the art	119
10.2	A simplified model of chromatin dynamics drives differentia- tion process in Boolean models of GRN	120
10.2.1	The model	121
10.2.2	Results	123
10.2.3	Conclusion	130
10.2.4	Future work	131
10.3	The effects of a simplified model of chromatin dynamics on attractors robustness in RBNs with self-loops	132
10.3.1	Experimental settings	132
10.3.2	Results	133
IV	Differentiation models in Robotics	139
11	Dynamical Systems Robotics	141
11.1	Dynamical systems in robot control design	141
12	On the design of robots with behavioural differentiation ca- pabilities	143
12.1	Attractor landscape as metaphor for designing robot controller	144
12.1.1	Attractors in robotics behavior	146
13	Biological Models in Robotics	149
13.1	GRN-based Models for Designing Robot Control	150
13.1.1	BN-based robot controller	154
13.2	GRN-based Models for Evolving Robot Morphology	155
13.3	GRN-based Models for Pattern Generation in Swarm Robotics	157
13.4	GRN-based Models for Co-Evolving Body and Brain	158
13.5	Morphogenetic Robotics	160

13.6 Conclusion	163
14 Robotics applications	165
14.1 Case study I – Attractor landscape: proofs of concept	165
14.1.1 BN model of a simple genetic network	165
14.1.2 Example 1: Controlled phototaxis	167
14.1.3 Example 2: Actions triggered by an external stimulus	170
14.1.4 Conclusion: Implications on robotics and synthetic bi- ology	173
14.2 Case study II: Engineering behavioural differentiation in BN controlled robots	173
14.2.1 Vision	173
14.2.2 Methodology	174
14.2.3 Results and Conclusion	175
14.3 Case study III: Online adaptiveness or adaptive semantics? .	176
14.3.1 Experimental setting	177
14.3.2 Results	179
14.3.3 Conclusion	180
V Conclusions and Future Perspectives	183
15 Ongoing and future work	185
Conclusion	189
Bibliography	208
List of Figures	212
List of Tables	213
Appendices	215
A The impact of self-loops on BN attractor landscape	217

Introduction

The behavioural and functional differences among biological cells are emergent and due to regulatory mechanisms which can turn on or off genes. Genetic regulatory networks (GRNs) epitomise the regulatory interactions among genes. Their dynamics drive, among other things, the biological cellular differentiation. Cell differentiation is the process whereby stem cells become progressively more specialised and by which a wide diversity of living cells and organisms, at the last, are produced. Dysregulation of cell differentiation is involved in many dangerous diseases, one out of all cancer. The comprehension of the processes underlying biological differentiation and a formulation of a general theory by which explaining, predicting and steering them are currently the main objectives of Systems Biology and Complex Systems Science. The work carried out in this dissertation aims to provide a contribution to the complex challenge of understanding differentiation, and at the same time, it tries to exploit its potential to design artificial agents able to attain not trivial tasks.

The approach adopted relies on the employment of Boolean networks-based computational models. Using computational models and *in silico* simulations, it is possible to explain biological results, make predictions and stimulate new hypotheses to be tested in wet labs.

In Figure 1 the whole thesis work is represented by a concept map. In the image, the original contributions proposed in this thesis are highlighted by transparent boxes.

In particular, the thesis presents an experimental validation study of theoretical results achievable by a recently proposed differentiation model. Then an automatic designing procedure for attaining Boolean networks able to express desired differentiation lineages dynamics is presented.

With the aim of capturing differentiation phenomena with higher accuracy, an in-depth study on the impact of the regulatory motif of *self-loops* has been performed. Self-loops have, indeed, so far been neglected in Boolean genetic networks, but they play a key role in the biological counterpart and therefore in differentiation.

Notably, a completely new model driven by an epigenetic mechanism, inspired by biological methylation, has been introduced.

The abstraction of these computational models finds natural application

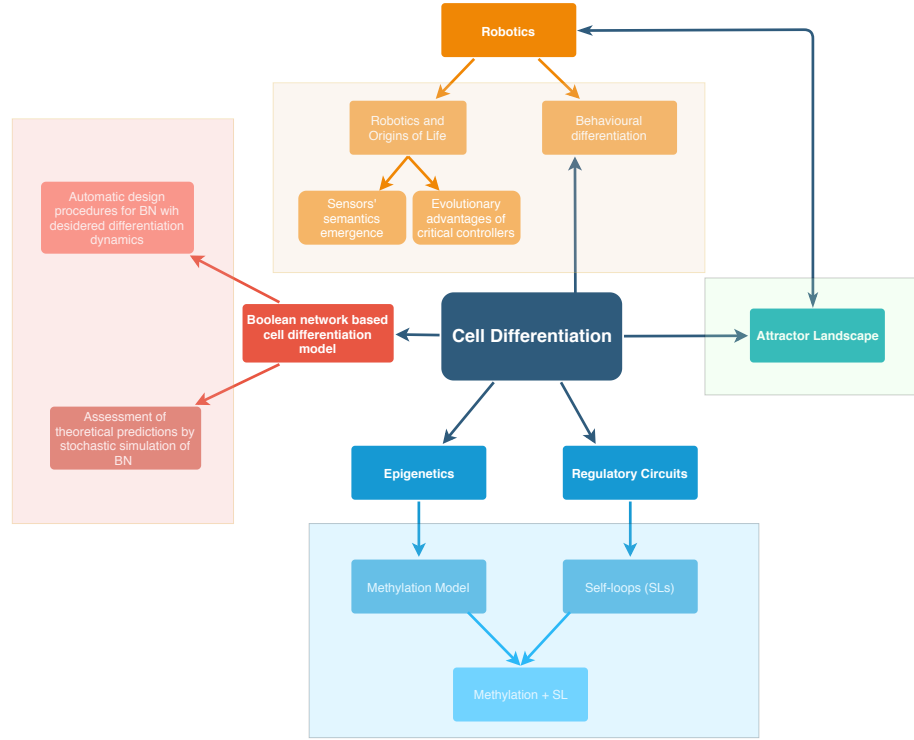


Figure 1: Concept map concerning the relationships between the original contributions produced by the thesis work.

where different behaviours are required, and robotics is a research field in which they can be profitably applied. A conceptual metaphor which takes inspiration from differentiation and which proposes itself as a new paradigmatic approach for designing robotic agents capable of expressing different behaviours is discussed. In support of this, some proofs of concept that support its potential are presented.

Thesis Contributions

Although all the scientific contributions will be thoroughly discussed in Part III and IV, the macro-objectives of this dissertation are here summarised in order to guide the reader.

The main contributions are the following:

- a critical analysis, highlighting advantages and limitations, of some aspects of current Boolean network-based computational models of differentiation;
- introduction of topological and dynamical network motifs able to reproduce some differentiation phenomenological characteristics;
- formulation of new (Boolean network-based) computational models that include—so far not addressed—relevant mechanisms underlying differentiation;
- implementation of software that represents abstractions of current differentiation models and enable networks simulation and data collection;
- designing of Boolean networks with desired differentiation lineages;
- designing of robotics agents capable of expressing different behaviours in response, in particular, to endogenous (dynamics driven) or exogenous (environmentally produced) stimuli.

Summarising, the thesis tries to achieve a more comprehensive understanding of properties and complexities arising from the process of differentiation by applying both analytical means—such as new computational models and statistical studies on ensembles of networks—and tools of synthesis—such as automatic design procedures for obtaining networks able to reproduce the differentiation lineages phenomenology in biological and artificial contexts.

The published works stemming from this research are presented in the following.

List of Publications

- [Braccini et al., 2017] Braccini, M., Roli, A., Villani, M., and Serra, R. (2017). *Automatic Design of Boolean Networks for Cell Differentiation*, pages 91–102. Springer International Publishing, Cham.
- [Braccini, 2017] Braccini, M. (2017). Applications of biological cell models in robotics. arXiv:1712.02303 To be submitted.
- [Braccini et al., 2018b] Braccini, M., Roli, A., Villani, M., and Serra, R. (2018b). A comparison between threshold ergodic sets and stochastic simulation of boolean networks for modelling cell differentiation. In Pelillo, M., Poli, I., Roli, A., Serra, R., Slanzi, D., and Villani, M., editors, *Artificial Life and Evolutionary Computation – 12th Italian Workshop, WIVACE 2017, Revised Selected Papers*, volume 830 of *CCIS*, pages 116–128. Springer.
- [Montagna et al., 2018] Montagna, S., Braccini, M., and Roli, A. (2018). The impact of self-loops in random boolean network dynamics: A simulation analysis. In Pelillo, M., Poli, I., Roli, A., Serra, R., Slanzi, D., and Villani, M., editors, *Artificial Life and Evolutionary Computation – 12th Italian Workshop, WIVACE 2017, Revised Selected Papers*, volume 830 of *CCIS*, pages 104–115. Springer.
- [Braccini et al., 2019a] Braccini, M., Montagna, S., and Roli, A. (2019a). Self-loops favour diversification and asymmetric transitions between attractors in boolean network models. In Cagnoni, S., Mordonini, M., Pecori, R., Roli, A., and Villani, M., editors, *Artificial Life and Evolutionary Computation*, pages 30–41, Cham. Springer International Publishing.
- [Roli and Braccini, 2018] Roli, A. and Braccini, M. (2018). Attractor landscape: A bridge between robotics and synthetic biology. *Complex Systems*, 27:229–248.

[Braccini et al., 2019b] Braccini, M., Roli, A., Villani, M., Montagna, S., and Serra, R. (2019b). A simplified model of chromatin dynamics drives differentiation process in boolean models of grn. *The 2019 Conference on Artificial Life*, (31):211–217.

[Braccini et al., 2019b] Braccini, M., Roli, A., Villani, M., Montagna, S., and Serra, R. (2019b). The effects of a simplified model of chromatin dynamics on attractors robustness in random boolean networks with self-loops: an experimental study. In *Proceedings of the XIV International Workshop on Artificial Life and Evolutionary Computation. Rende (CS), Italy, 18-20 September 2019*. To be published.

[Montagna et al., 2020] Montagna, S., Braccini, M., and Roli, A. (2020). The impact of self-loops on boolean networks attractor landscape and implications for cell differentiation modelling. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*. To be published.

Structure of the dissertation

The dissertation is logically organised into five (macro) parts, the contents of which are summarised in the following paragraphs.

Part I - Background and Motivations This section focuses mainly on exploring the biological context related to the process of cellular differentiation, which constitutes, at the same time, the background and the goal of the thesis.

Chapter 1 provides an overview of the molecular mechanisms underlying the production of protein—the complex molecules that govern the cell’s, tissue’s and organism’s functions—starting from DNA. Differential production of proteins is a result of the differentiation process—the subject of this scientific investigation. Chapter 2 synthesises the methods, techniques and modelling approaches applied for understanding cell differentiation phenomenon, highlighting their differences and therefore their intrinsic potential to answer the fundamental questions related to this process. The open questions, challenges and perspectives in the comprehension of cell differentiation are the subject of Chapter 3—which concludes this first part. Here, the objective of the thesis and the approach that was adopted to try to advance in understanding this is presented.

Part II - Dynamical systems view of Cell This part depicts the dynamical systems view of cell.

In this regard, in Chapter 4 a brief mathematical discussion of dynamical systems is introduced. Subsequently, in Chapter 5 the Boolean network model—a prominent discrete dynamical system model of genetic regulatory networks—used to investigate the differentiation process in the thesis is described. Eventually, Chapter 6 recapitulates the main works that contributed to the current conceptual—mathematically grounded—framework of cell differentiation as a dynamical system, whose attractors represent cell types.

Part III - Extensions to Current Models The content of this part—and the next one—is based on the scientific production reported on Section List of Publications. The work presented in these chapters took the models (based on Boolean networks) for cell differentiation as starting conditions.

The automatic procedure for designing Boolean network able to attain desired differentiation dynamics—described by means of tree-like structures—is introduced in Chapter 7. Chapter 8 reports the in-depth analysis of the impact of self-loops in Boolean networks models carried out during my Ph.D. studies. Although self-loops are important in the real genetic regulatory networks, their contribution to Boolean networks dynamics, especially in a

context of cell differentiation, had never been addressed in a systematic way before. Another fundamental mechanism contributes to the cell differentiation phenomenology, epigenetics. This is the subject of Chapter 10: here we introduce for the first time a methylation mechanism in Boolean network model capable of reproducing some essential properties of biological cell differentiation.

Part IV - Differentiation models in Robotics This part introduces the works related to the application of the differentiation process metaphor in the robotics field.

In Chapter 11 the idea—presented by Rolf Pfeifer in the book “Understanding Intelligence”—of using the dynamical systems approach not only for designing and analysing robot’s behaviours but also for the actual design of artificial agents is illustrated. Subsequently, Chapter 13 goes through the approaches presented in the literature about the use of genetic regulatory network models as robots controller. Finally, our contributions to the robotic field are presented in Chapter 14. Cell differentiation metaphor finds its natural application whenever an autonomous agent could take advantage of its ability to give rise to different (specialised) behaviours, the latter conditioned by environmental, external or internal signals.

Part V - Conclusions and Future Perspectives The dissertation concludes by reporting an overall evaluation of the whole thesis work. The evaluation is done by pointing out the strengths and weaknesses of the proposed approaches, methods, and models. Chapter 15—with the proposal of a future research agenda and the presentation of the ongoing work—begins to face the limitations and critical points that this whole research has encountered.

Part I

Background and Motivations

Chapter 1

Biological Cell

Here, the biological context on which this entire work is based will be briefly introduced. In this chapter, we provide an outline of the main concepts involved in biological cells (focusing mainly on eukaryotic cells) and, in particular, in the cell differentiation process, the main subject of this dissertation.

This description—made at a high level of abstraction—encompasses only the fundamental mechanisms of the biological cell that are relevant to the purpose of the work presented in later chapters, neglecting some others equally important details.

1.1 DNA

All living entities—organisms—are made of (and from) cells, marvellous chemical machineries able to perform specific complex functions and with the extraordinary ability to create copies of themselves [Alberts et al., 2013]. The higher types of organisms, such as humans, are the result of groups of cells, each of which can perform a specific function, appropriately arranged to create organisational levels—such as tissues—with emergent properties and functions not ascribable to the single cells level. What makes these emerging properties appear is given—in addition to the arrangement of the cells—by the interactions among the cells. In the following sections we will see that all of the incredible capabilities of which the cells—and ensembles of them—are capable of are the direct or indirect result of the differential production of proteins and of interactions among them. We, therefore, start to describe what represents the blueprint of the protein production, and so, of life, the DNA.

DNA (*deoxyribonucleic acid*) is a nucleic acid that encodes all the genetic information necessary for life. DNA is a long polymer¹ made from repeating units called nucleotides. DNA was first identified and isolated by Friedrich Miescher in 1869; at a later time, in the 1953 Francis Crick and

¹A polymer is a large molecule composed of many repeated subunits

James Watson, using experimental data collected by Rosalind Franklin and Maurice Wilkins, worked together to determine the structure of DNA at the University of Cambridge, England. Watson and Crick proposed that DNA is made up of two strands that are twisted around each other to form a double helix structure [Alberts et al., 2013].

The DNA is what of which genes are compound. Genes are the hereditary units that transmit information from parents to offspring (they contain the information necessary for the production of a protein or RNA). Within cells, DNA is organised into long structures called chromosomes. Each chromosome is composed of a very long DNA molecule along which hundreds or even thousands of genes are arranged. When a cell is preparing to divide, the DNA of its chromosomes is duplicated so that each daughter cell gets an identical set of genes. In each cell, the genes arranged along the DNA molecules encoding the information to build other molecules of the cell. In this way the DNA controls the development and maintenance of the whole organism [Campbell et al., 2008]. In Figure 1.1 we can see a schematic representation of a cell, the DNA, chromosomes and genes.

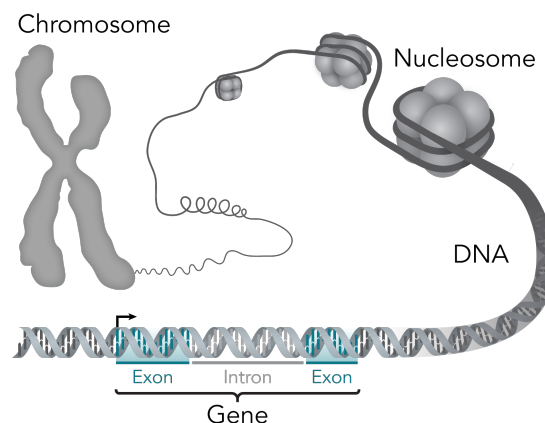


Figure 1.1: A schematic representation of a cell, DNA, chromosomes and genes. Image taken from <https://commons.wikimedia.org/wiki/File:Chromosome-DNA-gene.png>. Attribution: Thomas Splettstoesser (www.scistyle.com) [CC BY-SA 4.0 (<https://creativecommons.org/licenses/by-sa/4.0/>)]

1.1.1 From DNA to Protein

As we have seen in the previous section, the whole heritable genetic information of cells is preserved in the DNA (RNA in some viruses) and hence in the specific combinations of nucleic acids. To carry out the “instructions” coded into the DNA, cells transcribe it into RNA molecules, and afterwards,

translate these last into proteins, the functional means with which cells accomplish their tasks.

The actual process of translation from RNA to protein is carried out on ribosomes. Ribosomes are complex molecular machines and they are within all living cells; they serve as the site of biological protein synthesis. More precisely, the RNA molecule (mRNA), produced by transcription from DNA, is decoded by a ribosome to produce a specific amino acid chain or polypeptide. The polypeptide later folds into an active protein and performs its functions in the cell.

This process takes place only for those genes that belong to the coding DNA (nearly one per cent of the total DNA), and are called *protein-coding genes*, while the remaining part is called non-coding DNA and has mostly roles in epigenetic activity and regulatory interactions ².

The whole process by which the DNA instructions are converted into the functional product is called **gene expression**.

Francis Crick at the end of the 50s, synthesised all the flow of information that starts from DNA to RNA (transcription) and to the RNA to proteins (translation) with the well-know, and controversial, “central dogma of molecular biology”. The “dogma” explicitly rules out flows of information which concern protein-DNA, protein-RNA, protein-protein interactions ³: in this way it does not take into account fundamental processes as post-translational modification and epigenetics by now properly considered fundamental in the process of gene expression. It is precisely the limitations that derive from the conceptual framework depicted by the central dogma, but above all its incorrect declination in terms of one-to-one mapping from genotype to phenotype [Huang, 2012], that makes it now anachronistic and if not correctly understood deleterious for a correct understanding of the mechanisms that underlie the complex process of gene expression regulation.

1.2 Control of Gene Expression

All steps of gene expression can be modulated, since passage of the transcription of DNA to RNA, to the post-translational modification of the protein produced. Hence, gene expression is a complex process regulated at several stages in the synthesis of proteins.

1.2.1 Regulation of Gene Expression

In [Gilbert and Barresi, 2016] the stages responsible for the gene expression regulation are classified into four categories:

²https://en.wikipedia.org/wiki/Non-coding_DNA Date: 23/08/2019

³https://en.wikipedia.org/wiki/Central_dogma_of_molecular_biology Date: 23/08/2019

- *differential gene transcription* regulates the process of transcription which leads to the creation of nuclear RNA (nRNA) from genes;
- *selective nuclear RNA processing* determines which of the nRNA will enter the cytoplasm and become messenger RNAs (mRNA).
- *selective messenger RNA translation* modulates the process of translation from mRNA to proteins;
- *differential protein modification* concerns the several changes related to the post-translational regulation that determine whether or not a protein will be active.

Since in this discussion and in the remaining part of the dissertation we are interested in the *differential gene transcription* we will not consider the other, however important, processes.

Transcription factors (TF) are a particular kind of protein. They bind to specific DNA sequences in order to regulate the expression of a given gene. The power of transcription factors resides in their ability to activate and/or repress transcription of genes. The activation of a gene is also referred to *positive regulation*, while the *negative regulation* identifies the inhibition of the gene.

1.2.2 Gene Regulatory Networks

The entirety of intracellular regulatory interactions—among DNA, RNA, proteins (TF primarily) and other molecules [De Jong, 2002]—that are responsible for the up and down regulation of genes are epitomised in a complex structure termed *gene regulatory network (GRN)*, a term coined by Davidson’s group [Gilbert and Barresi, 2016].

The *regulation of gene expression* is essential for the cell, because it allows to control the internal and external functions of the cell. Furthermore, in multicellular organisms, gene regulation drives the processes of **cellular differentiation**, by leading the *differential gene expression*: the process that establishes the subset of genes that will be expressed. It is indeed the unique pattern of active genes that brings to the creation of different cell types.

In turn, the ultimate product of the differential gene expression is the different set of proteins that a cell can synthesise; these proteins have different ultrastructures that suit them to their functions and to the determination of the functions and the identity of the cell from which they are originating. Therefore, with few exceptions, all cells in an organism contain the same genetic material [De Jong, 2002], and hence the same genome (the haploid set of chromosomes of a cell). The difference between the cells are *emergent* and due to regulatory mechanisms which can turn on or off genes. It follows that two cells are different if they have different subsets of active genes.

It is precisely through the study of the complex interactions of gene regulatory networks and their consequent dynamics that we will study the properties of the process of differentiation, whose properties are shown in the following section.

1.2.3 Epigenetic mechanisms

Without entering into the historical and semantical debate of what we refer to with the term *Epigenetics*,—for a more thoroughly discussion to this regard see the work [Deans and Maggert, 2015]—we will utilize it with the only meaning of a further means to modulate access to genes and therefore to their regulation ⁴.

Eukaryotic cells are characterised by the organisation of DNA in a condensed structure, called *chromatin*. Chromatin is composed of nucleosomes, structures of DNA wrapped around octamers of histone proteins. *Histone methylation* and *histone acetylation* change—by adding methyl and acetyl groups to histones—the degree of compactness of the chromatin, in this way facilitating or obstructing gene expression.

Although methylation (acetylation) effects depend on the particular positions on histones on which it acts, it most often leads to tightly (loosely) packed regions of chromatin called *heterochromatin* (*euchromatin*) [Gilbert and Barresi, 2016, Perino and Veenstra, 2016, Schuettengruber and Cavalli, 2009]. These regions are not accessible neither by transcription factors nor by RNA polymerases and so the expression of genes belonging to these DNA areas is inhibited.

It is worth mentioning that methylation is tightly regulated by complex interactions, and that epigenetic dysregulation is very common in a lot of disorders, from cognitive, neurological and chronic diseases to cancer.

1.3 Cell Differentiation

Cell differentiation is the process whereby **stem cells** become progressively more specialised. The differentiation process occurs both during the development of a multicellular organism and during tissue repair and cell turnover in the adulthood. Gene expression, and therefore its regulatory mechanisms, plays a critical role in cell differentiation; as described in the previous section.

Stem cells are undifferentiated biological cells which can both reproduce themselves, *self-renewal* ability, and differentiate into specialised cells, *potency*. There are different kind of stem cells that can be classified according to their ability to generate progeny:

⁴See the supplementary material of [Huang, 2012] for a more detailed discussion on how epigenetics fits into the dynamical systems view of cell differentiation and development.

- *Totipotent stem cells* can differentiate into embryonic and extra-embryonic cell types and can construct a complete, viable organism. These cells are produced from the fusion of an egg and sperm cell. The fertilised egg, or *zygote*, is the ultimate totipotent cell because it has the capability to generate all the cell types of the body [Lodish et al., 2008];
- *Pluripotent stem cells* are the descendants of totipotent cells and has the capability of generating a number of different cell types, but not all. The cells of the embryoblast and the trophoblast are considered pluripotent, the former is the mass of cells inside the primordial embryo that will eventually give rise to the definitive structures of the fetus⁵ (they can form all types of cells that give rise to a human organism except those that form structures such as the placenta and other support membrane needed during gestation, also called *Embryonic stem cell*, ES in short [Hardin et al., 2012]), the second provide nutrients to the embryo and develop into a large part of the placenta⁶; see the Figure 1.2 for an example of differentiation tree, in this case related to hematopoiesis.
- *Multipotent stem cells* can differentiate in several cell types, but only those of a closely related family of cells⁷. The embryoblast gives rise to the three germ layers: the *endoderm* which forms the internal organ tissues (the stomach, the liver, the lungs etc.), the *mesoderm* which forms muscle, bone, circulatory system etc. and finally the *ectoderm* which ends up forming the skin and the nervous system⁸. The cells of the germ layers are multipotent stem cells. An example of multipotent stem cells are the Hematopoietic stem cells which may develop in different types of blood cells, hematopoiesis, but they can not develop into brain cells or other cell types outside of the types belonging to the blood tissue cells.
- *Unipotent stem cells*, found in adult tissues, divides to form a copy of itself plus a cell that can form only one cell type [Lodish et al., 2008], their own, but have the property of self-renewal, which distinguishes them from non-stem cells.

By means of the cell differentiation process cells acquire the *specialised properties* that distinguish different types of cells from each other. As cells acquire these specialised traits, they generally lose the capacity to divide. In a normal tissue, one of the two cells produced by each cell division retains

⁵https://en.wikipedia.org/wiki/Inner_cell_mass Date:25/02/2016

⁶<https://en.wikipedia.org/wiki/Trophoblast> Date:25/02/2016

⁷https://en.wikipedia.org/wiki/Stem_cell Date:25/02/2016

⁸<https://www.boundless.com/biology/textbooks/boundless-biology-textbook/gene-expression-16/regulating-gene-expression-in-cell-development-117/cellular-differentiation-464-13120/> Date:25/02/2016

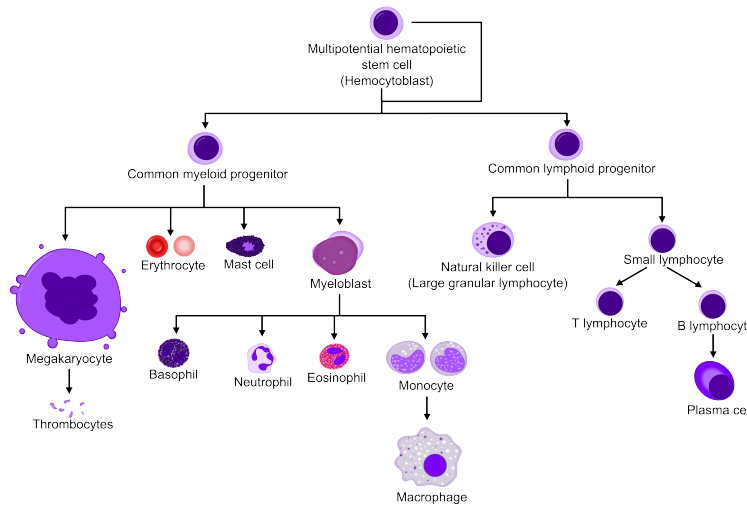


Figure 1.2: Example of differentiation tree of hematopoiesis. Image taken from https://commons.wikimedia.org/wiki/File:Hematopoiesis_simple.png. Attribution: Mikael Häggström (no attribution required), from original by A. Rad (requires attribution) [CC BY-SA 3.0 (<http://creativecommons.org/licenses/by-sa/3.0/>)]

the ability to divide, while the other cell loses the capacity to divide and finally dies; this ensures that there is no increase in the number of dividing cells. Cell division is therefore carefully balanced with cell differentiation and death. In *tumors*, this balanced arrangement is disrupted and cell division is uncoupled from cell differentiation and death. As result, some cell divisions give rise to two cells that both continue to divide, this process progressively increases the number of dividing cells. The tumor will grow because new cells are produced in greater numbers than needed, so the normal organisation and function of the tissue gradually become disrupted. Tumors are classified as benign or malignant: benign tumors grow confined in a local area and are rarely dangerous, whereas malignant tumors invade surrounding tissue. The term *cancer* refers to any malignant tumor [Hardin et al., 2012].

In 2006, Shinya Yamanaka, and his research team, succeeded to transform differentiated cells, from mice, into a pluripotent state. Such *induced pluripotent stem (iPS) cells* seem to have many of the same properties as Embryonic stem cells. To produce iPS cells Yamanaka forced cells to express four transcription factor proteins that are expressed by pluripotent cells. Since iPS cells are quite similar to ES cells might be ideal for medical treatments: they can be treated with various growth factors and they can be pushed to become various types of cells. In this way they might differentiate into nerve cells and then used to repair brain damage in patients with Parkinson or Alzheimer diseases or they might be utilised to replace

the defective pancreatic cells of patients with diabetes [Hardin et al., 2012].

1.3.1 Role of epigenetics in differentiation process

Biological cells exploit differential methylation to modulate their gene expression during development and differentiation. Remarkably, the attained configurations of DNA methylation are inherited and progressively extended as cells become more specialised [Kim and Costello, 2017]. Therefore, methylation contributes to maintain and stabilise the attained gene expressions that ultimately characterise the identities of the various cell states. In addition, the patterns of methylation can be inherited—see the genomic imprinting phenomenon for an example—and so influence not only the single cell behavior and the organism to which it belongs, but also to its progeny.

Chapter 2

Modelling Approaches to Cellular Processes

This chapter is devoted to the presentation of the most prominent approaches, often complementary and interdependent, applied for studying the dynamics of cellular processes.

As we will briefly see in the next sections, each of these approaches relies on the use of models. The latter, even when they remain implicit in the working assumptions, define the standpoints, the possible questions and so the methods and the technologies employable for investigating the biological phenomenon of interest.

After an informal presentation of the concept of model, we will go through the main approaches used in the study of complex biological phenomena, such as the differentiation process. The discussion of the principles and models, along with their benefits and limitations, used by these approaches will hopefully help the reader to frame the work and the scope of the attained results carried out during the doctorate and here reported. Indeed, the work here presented is mainly based on the modification and proposition of (new) models to investigate the processes underpinning cell differentiation by means of them.

2.1 What is a model?

I would like to start this discussion by citing some definition of what is a model:

[...] It is a model, something that does not really happen in nature, but which helps us to understand things that do happen in nature. Models can be very simple [...] and still be useful for understanding a point, or getting an idea. Simple models can be elaborated and gradually made more complex. If all goes well, as they get more complex they come to resemble the real world more.

Richard Dawkins, *The Selfish Gene* [Dawkins, 1976]

A model is a system we decide to use to represent another system.

Roberto Serra, personal communication.

Remember that all models are wrong; the practical question is how wrong do they have to be to not be useful.

George Box, *Empirical Model-Building and Response Surfaces* [Box and Draper, 1987]

No substantial part of the universe is so simple that it can be grasped and controlled without abstraction. Abstraction consists in replacing the part of the universe under consideration by a model of similar but simpler structure. Models, formal or intellectual on the one hand, or material on the other, are thus a central necessity of scientific procedure.

Arturo Rosenblueth and Norbert Wiener,
The Role of Models in Science [Rosenblueth and Wiener, 1945]

What emerges from these informal definitions is that a model is an abstract and schematic representation of a system. A model emphasises only a portion of the system, it captures only some of its features. The formulation of a model requires an abstraction process, which involves simplification, aggregation and omission of details. Therefore, the definition of a model, like that of a system itself, requires to say what belongs to the model and what does not, which usually depends on the observer's point of view.

A model allows us to:

- understand and investigate some properties of a system;
- control and steer it;
- make predictions on its future.

In “hard” sciences, models are often formal mathematical representations of systems; however, they can be conceptual, verbal, diagrammatic, physical, etc [Sayama, 2015].

In the following sections, we summarise the ideas, methods and models underlying different approaches applied to the study of cell differentiation or more in general to cellular behaviours.

2.2 Molecular biology approach

Molecular biology is a branch of biology, and as the name clearly suggests it was from the beginning (around the 30s) concerned to the molecular basis of cell’s functions and activities ¹. In detail, it refers to the understanding of the molecular basis of the process of replication, transcription and translation of genetic material, i.e. genes ².

Some sentences by William Astbury [Astbury, 1961] well synthesise the purposes and needs that led to the birth of molecular biology:

[...] not so much a technique as an approach, an approach from the viewpoint of the so-called basic sciences with the leading idea of searching below the large-scale manifestations of classical biology for the corresponding molecular plan. It is concerned particularly with the forms of biological molecules and [...] is predominantly three-dimensional and structural—which does not mean, however, that it is merely a refinement of morphology. It must at the same time inquire into genesis and function

So, the molecular level understanding sought by molecular biology is—first of all—performed by identifying and reporting all the molecules or groups of molecules that participate in a given phenomenon. The focus is therefore mainly on the **structure** of the interactions between molecules that give rise to phenomena such as cell cycle state change, fate change, etc. Furthermore, the research often involves only **isolated parts** of a cell or organism.

In addition to the discovery of chains of regulatory interactions between genes, the major efforts of molecular biology concern the identification of *signal transduction pathways*. They are cascades of chemical reactions, within cells, that occurs when a molecule, ligand, attaches to receptors on the cell membrane. The cell’s response to this signalling cascade is a change in the biological activity. Signal transduction pathway are of fundamental relevance in molecular biology since aberrant activities of these may result in diseases. For this reason, they are also the target for drug therapy in disease conditions.

¹https://en.wikipedia.org/wiki/Molecular_biology Date: 08/10/2019

²Obviously, there are overlaps and no clear distinctions between molecular biology and other fields of biology, such as chemistry, genetics and biochemistry.

2.2.1 The end of central dogma?

The entire chain of information, along with its transformation, leading from genes to proteins, described in the Section 1.1.1, and known as “central dogma of molecular biology”, has represented for years, albeit in a simplified way, the goal of molecular biology.

In the last years, in light of the emerging results concerning the roles of RNA and more in general of protein-to-protein interactions in regulation, central dogma has undergone some criticisms and revisions. Indeed, a broader (mis)interpretation of the *dogma* has fostered the idea that all observable cellular changes (cell cycle, phenotype, drug response, etc.) were a direct result of the action of **independent** signal transduction pathways. This linear causal relationship between single molecules, or signalling cascade of these, and cellular phenotypic manifestations ³ has been challenged by the observations of cross-talks between many signalling pathways and the multiple effects caused by the same cascade (see WNT pathway). But what has more undermined the classical view of biology has been the advent of systems biology and the notion of (complex) network of regulatory interactions between molecules. Since then, the network concept, and the related promising results, has been superseding the (ad-hoc) explanations of cellular characteristics based on superposition of independent signalling cascades [Zhou et al., 2014].

2.3 Systems Biology approach

Systems biology is an interdisciplinary field of study which studies living organisms as systems that evolve over time [Ideker et al., 2001, Aderem, 2005]. “The whole is greater than the sum of the parts” is the most used expression to present Systems Biology’s vision. This catchy phrase tries to communicate the change in the approach determined by systems biology in the field of biology research. The change mentioned concerns the transition from the more traditional reductionist paradigm typical of molecular biology to the so-called holistic approach: a **system-level understanding of biology** [Kitano et al., 2001].

The system-level understanding advocated by systems biology enlarges the *modus operandi* classically pursued by molecular biology. Indeed, as we have seen in the previous section, molecular biology is mainly devoted to identifying all the genes and proteins and their **static interaction relationships** involved in a particular cell’s or organism’s function. While systems biology tries to identify how the various components, and components of components, **dynamically interact** [Kitano, 2002]. For reaching this broader comprehension, systems biology uses a combination of technolo-

³In other words, between genotype and phenotype.

gies, from high-throughput precision measurements to sophisticated computational tools and analysis. But it is on the construction of models and the possibility of hypothesis-based research—which is enabled by these—on which systems biology leverages.

In this framework, systems biology guides the development of new models, technologies and computational tools and at the same time, in turn, these revolutionise biology. This conceptual schema is well depicted in the Figures 2.1: the global observations are matched against model predictions (hypotheses) in an iterative manner, leading to the formation of new models, new predictions, and new experiments to test them.

For what concerns the studies of cell's processes, systems biology focuses mainly on the complex structure of interactions and feedbacks between genes and proteins and its emergent and general properties, structure represented in Gene Regulatory Networks' models. This approach has been called “Complex Systems Biology” by Kunihiro Kaneko [Kaneko, 2006].

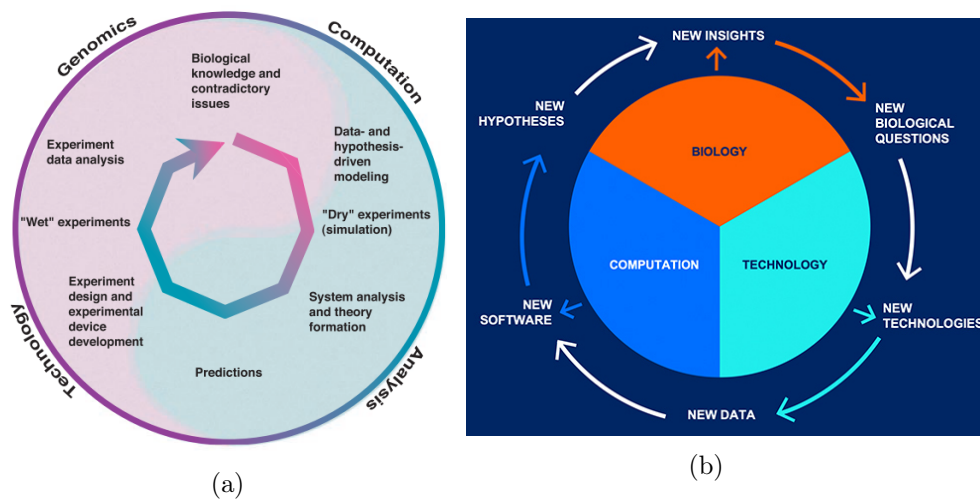


Figure 2.1: a: Depiction of the systems biology approach. Image taken from [Kitano, 2002]. b: Different representation of the systems biology approach. Image taken from <https://isbscience.org/about/what-is-systems-biology/>; all credits go to *Institute for Systems Biology (Seattle, USA)*.

In conclusion, molecular biology and systems biology are two different—but complementary—approaches. No understanding at the system level is possible without knowing in detail the parts that compose it. Vice versa, the detailed structure of a component cannot be separated from a behavioural analysis that characterises it not only at the organisational level to which it belongs but within the entire organism, the latter result of the organisation in interacting levels of increasing complexity capable of revealing emergent

properties⁴. Therefore, only when both the two approaches are applied a higher level—and more complete—understanding can be achieved. [Stevens, 2004]

2.4 Complex Systems Science approach

Complex systems science (CSS) is the ensemble of theories and methods that deals with complex systems: the study, modelling and control of these systems embody its purpose. Examples of complex systems are the brain, the society, the universe, the cell.

CSS is interdisciplinary and it involves methods and tools from mathematics, physics, computer science, biology, economy, philosophy, neurology and more. Although there is no formal definition of CSS, we can certainly recognise the properties exhibited by the systems it is interested in. The systems taken into account by this research field exhibit some of the following properties:

- composed of many elements;
- nonlinear interactions;
- non-trivial network topology;
- positive and negative feedbacks;
- adaptiveness;
- evolvability;
- robustness;
- multiple levels of organisation;
- self-organisation;
- emergence;
- universality.

CSS tries to understand the causes of the properties listed above, properties that are often not observable in the parts that make up the system. Hence, this field of science studies how parts of a system give rise to the collective behaviours of the system, and how the system interacts with its environment. It focuses on certain questions about parts, wholes and relationships.

⁴Emergent properties are properties not ascribable to individual parts of a system and that cannot be predicted even with full understanding of the parts.

From these considerations, we can see how the aims of CSS resemble those of systems biology; indeed, systems biology can be seen as a subset of the complex systems science applied to the understanding of biology. This can be appreciated in the Figure 2.2, which depicts the organisational map of CSS in seven topical areas.

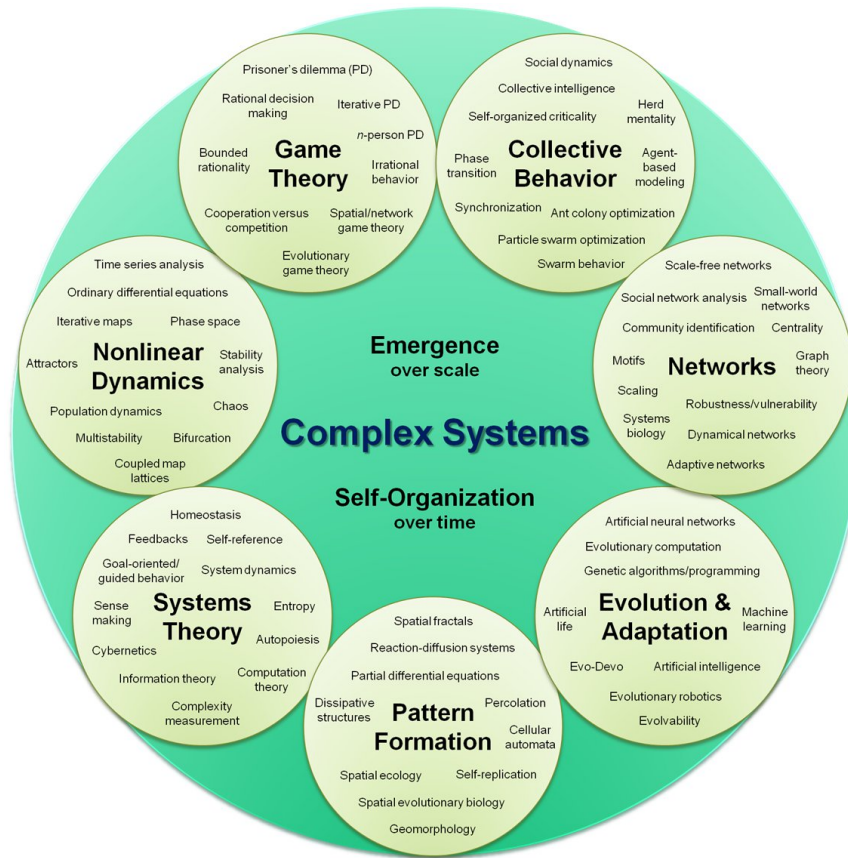


Figure 2.2: Organisational map of complex systems science, image taken from [Sayama, 2015].

The main issues addressed by CSS can thus be summarised by these questions:

- “how do parts of a system give rise to its observable **collective behaviours**?”
- “how does the system **interacts** with its environment?”
- “what are—and where reside—the sources of **order** and **self-organisation**?”

As regards the study of biological processes and cellular behaviours (to which cellular differentiation belongs), CSS provides valid theoretical and experimental tools to address the many complexities they present. Two of the

most prominent classes of approaches applied for the study of cellular processes and belonging to CSS are those of *dynamical systems* (see Chapter 4 for a more in-depth discussion) and of *network theory* [Barabási et al., 2016]. By way of example, *cellular automata* [Von Neumann et al., 1966], *Boolean networks* [Kauffman, 1969b], but also *differential or difference equations* for the first and *threshold networks* for the second have proven to reproduce some important cellular phenomena.

In this dissertation, we will see applied various methods and concepts coming from the typical approaches of CSS. In particular, the use of **non-linear dynamic system models** together with the search for **emergent behaviours** and **general properties** deriving from the **dynamics of networks** of many entities will be the means we use to investigate the complexities underlying the differentiation process. To be more precise, all the investigations carried out, aimed at investigating the root causes of the differentiation process, are based on Boolean networks, well-known and prominent models of gene regulatory networks. They will be described in detail in Chapter 5. What led to the choice of this model is their ability to well represent generic properties of differentiation without resorting to fine-tuning or fitting techniques in models construction [Geard and Willadsen, 2009]. In addition, their Boolean nature avoids the need to maintain a high number of kinetic parameters, otherwise necessary and computationally expensive in simulations of continuous models of gene regulation process.

2.5 Computational approaches

In this section, we will briefly present the computational approaches to the modelling, description and understanding of complex systems, such as biological ones.

Examples of computational approaches are *rule-based systems*, *rewriting systems*, *process algebras*, *agent-based systems* and *connectionist approaches*. **Agent-based modelling** (ABM), in particular, plays an important and recognized role in the modelling of complex and biological systems [Montagna and Omicini, 2017]. The agent definition is subject of debate within the scientific community that deals with ABM. For the purposes of this discussion, we can define them as autonomous, proactive entities, capable of perceiving and acting in an environment.

ABM are computational models that exploit the simulation of interacting agents in order to investigate the relationship between the micro-scale (agents') and macro-scale (whole system's) levels' properties. Therefore, in the same way as Complex Systems Science and Systems Biology, ABM's inquiries are concerned with **emergence**: how system properties appear from system's components interactions.

The employment of ABM approaches, or any computer-based approaches

in general, shows advantages over equation-based ones (these last typically of Complex Systems Science methods) when we are in the following cases [Serra and Villani, 2006]:

- the elements of which the model is composed must perform sophisticated information processing capabilities;
- the entities must possess an internal structure;
- the heterogeneity among the entities—required by the problem—cannot be resolved by choosing different parameters for the equations.

ABM approaches—and above all its multi-agent declination, *multi-agent system* (MAS)—have been successfully applied for simulating multicellular systems, i.e. organisms composed of numerous interacting cells. Here, an agent often is the modelling counterpart of the cell. In literature we can find agent-based models capable of reproducing *in silico* some spatial and temporal structures (e.g. pattern formation) of different biological phenomena, like *Drosophila melanogaster* morphogenetic process [Montagna et al., 2010b, Montagna et al., 2010a], cancer growth and invasion [Wang et al., 2015], but also for optimising cancer cell's reprogramming [M Biava et al., 2011] and others [Zhang et al., 2009].

To the best of our knowledge, no agent-based approach or other computational techniques have begun to address the challenges of understanding the whole differentiation process.

Chapter 3

Cell Differentiation: challenges and complexities

After presenting what represents the biological background and the main modelling approaches for cell differentiation, in this chapter, we clarify the objective of the thesis.

3.1 Open questions and perspectives

Although many molecular mechanisms (signal transduction pathways, epigenetic processes, network motifs and others) which enable the cells to express different proteins are known, what is missing is an understanding of how cells *orchestrate* and *maintain* these mechanisms in a robust way in order to ultimately determine different cell types, each one with a defined function. Besides, what is missing is a **general theoretical framework** to describe the phenomena involved in cell differentiation. In this framework, it would be possible to insert and contextualise the current—and future—knowledge concerning cell differentiation.

Today the whole complex of knowledge related to differentiation is summarised by some *principles* and *observations*, results of *wet* or *in vitro* experiments, which have been consolidated and accepted as such over time. But these are sometimes in opposition to each other, or together, they depict an incomplete context: thus forcing the experimenter or scientist to create ad-hoc explanations each time that the accepted principles do not explain a new experiment's result. In conclusion, the lack of a theory prevents hypothesis-driven experiments.

In this regard, for example, we report the following fallacy which points out the need for a formulation of a general theory of differentiation. Although the abundance of the phenotypes that make up multicellular organisms is the result of the very same genetic code, DNA, why every malignant manifestation (whether this is a cancer cell type or not) is attributed—by molecular

biology—to genetic mutations ¹?

Many approaches belonging to those presented in Chapter 2 could represent valid means to investigate the problems and questions, still unsolved, related to differentiation process. In this thesis, we have pursued research in understanding biological differentiation by applying a model-based approach. However, various model-based approach and various declinations of the concept of model exist in science [Piscopo, 2013]. To a large extent, the discussions about which models are most appropriate for a problem concern the philosophy of science—a treatment of the latter goes beyond the scope of this dissertation. So we just summarise some critical aspects of the modelling possibilities in relation to a given biological phenomenon and contextualise our work in light of this.

In the Figure 3.1, we highlight the possible different modelling choices, their mutual relations and their links with the target real system. The specific example in the above-mentioned figure—generalisable to any other target system—is biologically inspired by a population of cells composed of two subpopulations each expressing a characteristic protein which distinguishes it from the other. The distribution of the number of cells presenting each of the two proteins is reflected in the “green level”: the so-called **data model** since it could be obtained by applying statistical methods. Instead, the lowest level contains the “continuum” of modeling approaches capable of giving rise to the statistics represented in the green level, characteristics of the real target system under consideration. Here, without going into fine-grain details, we can surely identify two antipodal approaches ²:

mechanistic models: those models which strongly rely on the one-to-one relation between model’s components and system’s ones, i.e. a sort of isomorphic relationship between the model and its target system. Please note, as Giere [Giere, 2004] points out, it is not the model itself that creates the representational binding between the real’s and model’s entities, rather it is the modeller—who owns the similarities between the two systems in mind—which creates this link by projecting a meaning/representation on the entities of the model.

statistical models: on the contrary, statistical models are those with no apparent (or very loose) links between model and target system—in some ways we can also consider these models as data models.

Following the above distinction, our approach is based on the use of models of mechanistic type. Indeed, what will be presented in the following chapters relies on a well-know model of genetic regulatory networks, namely

¹[Huang et al., 2009b, Huang, 2012]

²This classification takes inspiration from—and to some extent reflects—the distinction between *open-box* and *closed-box* problems made by Wiener in [Rosenblueth and Wiener, 1945]

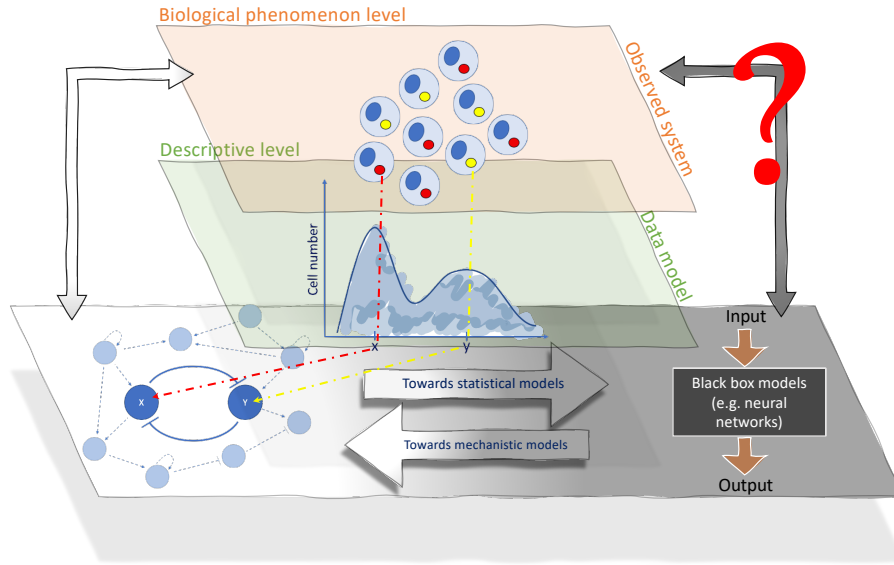


Figure 3.1: With this pictorial drawing, we synthesise the possible modelling approaches, their mutual relationship and the link with the target system. In particular, the red level describes a population of cells composed of two subpopulations, each expressing a specific protein (red or yellow in the image). The green level represents statistics of the number of cells expressing the two proteins. Eventually, the lowest level tries to portray the continuum of models that can give rise to the statistics mentioned above.

Boolean networks (BNs). The use of BNs is motivated by their capacity to express relevant properties of real genetic networks ³, but also because their nodes represent the genes of the target biological organism (or class of organisms). By using them, we exploit the fact that the cause and effect relationships found in these models can be reported/found, with due limitations, to the actual modelled system. Therefore, they try to go beyond the simple correlations and to investigate what could be the causalities between the components of a system and their effects. The use of models and the related hypotheses (obtained through simulations if they are computational models) represent a step towards a formulation of a general theory. A theory

³See Chapter 5 for a more for a more in-depth discussion of their properties.

that can produce verifiable predictions, and therefore be falsified or revised if these last are not in accordance with observations.

The use of dynamical systems as a theoretical framework and of computational models based on Boolean networks as a means to put in place its principles represent a step towards a general theory of the cell differentiation process.

In this thesis, the computational models used for investigating biological differentiation properties are also used for the construction of artificial complex systems. Indeed, these biologically-inspired models are used as control software in robotics. This is firstly motivated by the fascinating dynamics and wealth of possibilities offered by the biological differentiation process. Secondly, this approach can exploit the analogy among fully differentiated cell types and robotic behaviours—in a context where these last can be the emergent results of the coupling between robot and environment. This could cope with the necessity, from the robotic perspective, to overcome the traditional paradigms for designing robot controller, these last unable to reproduce very different behaviours with the same controller instance. The differentiation metaphor can, therefore, represent a valid candidate to attain more complex robot behaviours. Here, we start exploring this possibility by introducing some fundamental abstractions for designing robots with behavioural differentiation capacities and some proof of concepts of their applicability.

Part II

Dynamical systems view of Cell

Chapter 4

Dynamical Systems

We have seen in the previous Chapter 2 that the dynamical systems approach, combined with methods and technologies deriving from Systems Biology, represents one of the most promising means to investigate and reproduce cellular dynamics. Therefore, this chapter offers a glimpse of what underlies dynamic systems, but it will not be a complete and formal treatment of the theory of dynamic systems. It aims at elucidating the fundamental mathematical concepts that underlie all the research reported in this dissertation, thus providing a framework and a common glossary for the remaining part of the thesis.

Dynamical systems theory is the branch of mathematics that deals with systems that evolve in time. Indeed, a **dynamical system** is a system that it is described by means of a rule that govern the time evolution of its *state*. The time-evolution rule is usually formally defined by means of **differential equations** and **difference equations** (also known as iterated maps). By analysing differential equations, that are the most used in physics and engineering, we can distinguish them in ordinary or partial differential equations, depending on whether there is only one independent variable or more than one. A general formulation of an n th-order ordinary differential equation (ODE) is the following one:

$$\mathbf{F}(t, \mathbf{X}(t), \mathbf{X}'(t), \dots, \mathbf{X}^{(n-1)}(t)) = \mathbf{X}^{(n)}(t) \quad (4.1)$$

The Equation 4.1 contains an explicit time dependency and for this reason it is said to represent *nonautonomous* systems, contrary to the so-called *autonomous* systems which does not present it. We can always remove a time dependence by adding an extra dimension to the system [Strogatz, 2018], and transform it into a $(n+1)$ -dimensional system. In addition we can always, by introducing new variables, reduce the Equation 4.1 into this one:

$$\mathbf{F}(t, \mathbf{Y}) = \frac{d\mathbf{Y}}{dt} \quad (4.2)$$

Differential equations that cannot be written as a linear combination of their derivatives are called *nonlinear* systems. Nonlinear systems, although often difficult to solve analytically, are very interesting because many natural phenomena cannot be represented solely by linear functions. So very often they are solved by performing approximations or relying on numerical computations. Cellular dynamics and above all gene regulation are also governed by strong non-linearities.

The number of *degrees of freedom* [Hilborn et al., 2000] of a dynamical system is the number of independent variables needed to describe its *states*. The set of all possible states is called *state space*. A dynamical system by starting at an initial condition (i.e. a state of the state space) and by following the time evolution rule describes a *trajectory*, namely a sequence of states. If all functions, that compose the left-hand side of the Equation 4.1, are continuous and at least once differentiable (Lipschitz condition) then only one solution can pass through a given point in the state space, and so two different trajectories cannot intersect [Hilborn et al., 2000].

4.1 Attractor concept

The geometric figures in the state space that attract a number of distinct trajectories are called **attractors**. The set of initial conditions that end in a specific attractor is called **basin of attraction** for that attractor.

Attractors can be divided into three categories: **fixed points**, **limit cycles** and **strange attractors**¹ Here the first two classes will be discussed, while the last one will be introduced later.

A *fixed point* (or equilibrium point or steady state) is a point in the state space for which all time derivatives of the state variables are 0:

$$F(t, Y) = 0 \quad (4.3)$$

From Equation 4.3 it is apparent that they are so-called equilibrium points because they remain constant for all time.

$$X(t) = X^* \text{ if } X = X^* \text{ initially} \quad (4.4)$$

It is important to note that solutions of the Equation 4.3 are not necessarily attractors: to be called attractors they have to be **stable** equilibrium points. Instead, a *limit cycle* is a periodic trajectory that is isolated from the neighbouring trajectories. Also limit cycles are attractors if they are stable limit cycle.

¹Although the degrees of freedom influence the possible long-term behaviours of a system, all the conditions in which they can appear, and their related properties, go beyond the scope of this discussion.

4.2 Stability

Often when we are dealing with complex non-linear systems there are no analytical solutions to the equations that describe them. This means that we cannot obtain all the fine-details regarding the time behaviour described by the system [Serra and Villani, 2006]. However, it is possible in these cases to make use of some theorems that give us information about their asymptotic dynamics—i.e. when time approaches infinity [Kaplan and Glass, 2012]. Only the theorems and methods related to the **stability** of equilibrium points will be here presented since from a theoretical point of view they play an important role for the scope of our discussion and for the whole thesis work.

Stability theory study the conditions under which solutions of dynamical systems are stable (i.e. trajectories remains in the vicinity of the solution) under small perturbation ². Stability can be characterised in the following ways [Nicolis and Prigogine, 1989]—supposing that \mathbf{X}_s represents the equilibrium point (that is, it verifies the Equation 4.3):

Lyapounov stability if for every ϵ we can find a perturbation smaller than δ — $\|\mathbf{X}(t_0) - \mathbf{X}_s\| < \delta$ —such that we have for every $t \geq t_0$ we have $\|\mathbf{X}(t) - \mathbf{X}_s\| < \epsilon$;

Asymptotic stability if it is Lyapunov stable and we can find a perturbation smaller than δ such that $\lim_{t \rightarrow \infty} \|\mathbf{X}(t) - \mathbf{X}_s\| = 0$

It follows that we call **unstable** any solution with trajectories that—under small perturbations—move away from it.

Non-linear systems express different behaviours for different values of their parameters. **Bifurcation theory** studies the situations in which a system can exhibit qualitative changes with respect to the variation of some control parameters. A bifurcation corresponds to the appearance or disappearance of new steady states and the change of their stability. Bifurcations can occur in both continuous systems and discrete systems.

4.3 Dynamical criticality

In *complex system science* there is a long-standing conjecture—i.e. the *criticality hypothesis*—which states that systems in a dynamical regime between order and chaos show behaviours characterised by an optimal balance between robustness and adaptiveness and with the highest computational capabilities [Roli et al., 2018, Kauffman, 1993, Kauffman, 1996, Packard, 1988, Langton, 1990, Crutchfield and Young, 1990, Prokopenko, 2013, Aldana et al., 2007]. In the study of complex living systems, this conjecture led Packard and Langton [Kauffman, 1993] to the formulation of the famous

²https://en.wikipedia.org/wiki/Stability_theory Date: 19/10/2019

expression “*life exists at the edge of chaos*”. The notion of criticality, more precisely of *critical state*, stems from the study of thermodynamic systems: it describes an abrupt change in certain system’s macroscopic properties as some control parameters (temperature, pressure, ...) are changed. Systems show interesting properties at the *critical point* of their control parameters [Roli et al., 2018]:

- despite microscopic differences between systems the same (macro) behaviour can be observed (*universality*);
- information exchange between distant parts of the system is maximal;
- no characteristic scale of response to perturbation exists (*power law*);
- long time to absorb perturbations (*critical slowing down*).

A system is therefore defined *critical* if it has one of the properties mentioned above.

To understand the meaning of criticality in dynamical systems—*dynamical criticality*—and the relevance that the *criticality hypothesis* conjecture has in the study of complex dynamical systems—and therefore also in fields such as biology, computer science and many others—it is necessary to briefly introduce the concepts of ordered and chaotic dynamical regimes and their main characteristics.

Ordered and chaotic regimes The qualitative behaviours of dynamical systems vary by acting on the parameters on which they depend. Systems that exhibit common characteristics are classified into what are called dynamical regimes. There are three dynamical regimes: the *ordered* one, the *disordered* (or *chaotic*) one and the *critical* one. Thus the parameter space is made up of three kinds of regions, which induce very different systems behaviours.

A dynamical system in an *ordered regime* shows steady states characterised by regular patterns, fixed points or limit cycles. Furthermore, systems in order regime are robust against perturbations: perturbation does not spread throughout the system.

On the contrary, systems in chaotic regime show aperiodic behaviours, apparently random in nature, as those which own **strange attractors**³. In addition, in systems in a disordered regime, perturbations tend to grow, causing close trajectories to diverge very fast. These systems are very sensitive to initial conditions, making long-term prediction very difficult, if not impossible.

Systems with parameters that lie between the ordered the disordered ones are called *critical systems* and they show properties that are a mixture

³Attractors with a fractal structure.

of those present in the other two regimes. In these systems, instead of dying out or growing, perturbations on average remain constant over time.

Studies that have used information theory measures have brought to light evidences that critical systems, at the phase transition, are those with the highest computational capabilities. These evidences have suggested a reformulation of the criticality hypothesis in “*computation at the edge of chaos*”. On the other hand, works based mostly on Boolean models of gene regulation networks have suggested that critical—or slightly sub-critical—biological systems best represent cellular dynamics [Kauffman, 1993, Shmulevich et al., 2005, Villani et al., 2018, Serra et al., 2007]. Noteworthy, some works suggest that also the human brain is dynamically critical [Beggs and Timme, 2012].

This is just a, far from comprehensive, list of the works on critical hypothesis and attempts to prove it. They highlight its importance for understanding real systems and also for designing artificial ones with capabilities comparable to these.

For a more detailed survey, we refer the reader to this remarkable and exhaustive study about dynamical criticality [Roli et al., 2018].

4.4 Robustness

In this discussion, we will often refer to the concept of robustness. Therefore, a definition of it is necessary to avoid future confusions, which may especially occur in the distinction between stability and robustness. In this section, we will try to clarify the differences between these two related, but different, concepts.

According to [Jen, 2003], *robustness* is a broader concept than stability—which as we have seen in the previous sections is a well-defined mathematical notion in dynamical systems. In a nutshell, it is related to *feature persistence* under a wider spectrum of perturbations, possibly multiple, of different nature (topological, environmental, etc.) in systems, or for systems’ features, that are difficult to quantify with metrics.

In the remaining, by the expression “attractor robustness” we mean the metric that synthesises the stability of attractors—probability of returning to the very same attractor—under the influence of a multitude of perturbations events. This definition is in accordance with the definition of “robust adaptation”, which Kitano gives in his work *Biological Robustness* [Kitano, 2004].

Chapter 5

Boolean Networks

Boolean networks (BNs) are a prominent example of complex dynamical systems presented by Kauffman in the late sixties [Kauffman, 1969b] as a genetic regulatory network (GRN) model. Since its introduction, they have proved capable of reproducing relevant phenomena in gene regulation [Kauffman, 1969b, Nykter et al., 2008, Shmulevich et al., 2005, Yuan et al., 2016, Su et al., 2017, Huang and Ingber, 2000, Helikar et al., 2012]. However, Boolean networks are not interesting only for biological modelling, but—as partly we will see in Part IV they also offer intriguing perspectives in engineering and computational contexts, given they can exhibit rich and complex behaviours in spite of the compactness of their description.

From a formal mathematical perspective, a BN is a discrete-state and discrete-time dynamical system whose structure is defined by a directed graph of N nodes, each associated to a Boolean variable x_i , $i = 1, \dots, N$, and a Boolean function $f_i(x_{i_1}, \dots, x_{i_{K_i}})$, where K_i is the number of inputs of node i . The state of the system at time t , $t \in \mathbb{N}$, is a N -tuple in $\{0, 1\}^N$, (x_1, \dots, x_N) , defined by the array of the N Boolean variable values at time t . Boolean networks are a generalisation of Cellular Automata (CA) [Von Neumann et al., 1966], indeed nodes can have different updating functions and these can be influenced by cells that are not necessarily in their spatial neighbourhood.

A simple example of BN for illustrative purposes is showed in Figure 5.1.

The most studied BN models are characterised by synchronous dynamics—i.e. nodes update their states at the same instant—and deterministic functions¹. The *state space* is finite— 2^N possible states for a BN with N nodes. Under a synchronous and deterministic dynamics every state has only one successor and therefore starting from any initial condition after a certain number of steps eventually a sequence of states will be repeated. Such se-

¹In the remainder of the work, we will refer to these configurations if not otherwise stated.

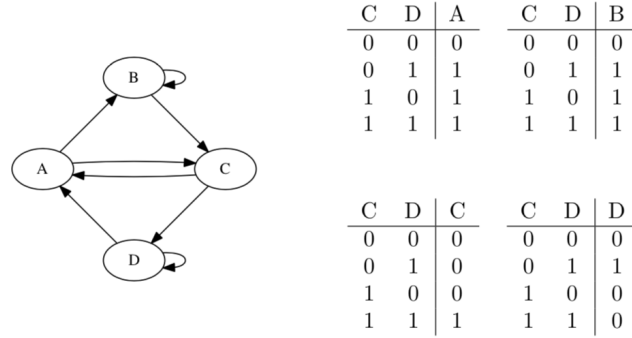


Figure 5.1: Example of Boolean network with its relative set of Boolean functions associated to its nodes, taken from [Roli and Braccini, 2018].

quences are called *cycles* or *attractors*: more precisely *cyclic attractors* if their period lasts more than one step, *fixed points* their special cases where the period is equal to one. An example of Boolean network with two cyclic attractors is reported in Figure 5.2. In accordance with dynamical systems, the succession of states traversed by the BN is called *trajectory* and the set of states that leads towards an attractor is called *basin of attraction*.

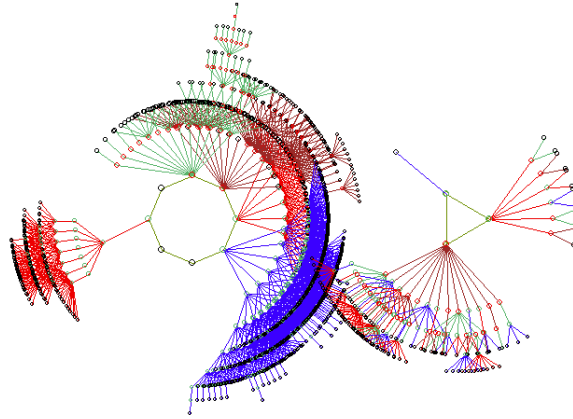


Figure 5.2: Example of BN's attractors computed and represented with the DDlab software [Wuensche, 1996]. In the graphical representation typical of DDlab each basin of attraction is depicted starting from the outside and going inwards, thus having the garden-of-Eden states at its extremities, if any, and the states of which the attractor is composed in the centre.

Despite their simplifications, they proved to be suitable systems to represent the dynamics of biological GRNs to many levels of abstractions [Graudenzi et al., 2011, Serra et al., 2006, Serra et al., 2007, Shmulevich et al., 2005]. In addition, given their level of abstraction, they can reproduce dy-

namics of systems belonging to very different domains, from natural to artificial or social ones. Historically, however, they have fulfilled a very important role in the modelling of regulatory interactions among genes. In their graph description, when we adopt them as GRN models, there is an arc from a node to another one if the expression of this last is influenced by the product (protein) of the previous one.

5.1 Random Boolean Networks (RBN)

In its original—and most studied—formulation called random Boolean network (RBN) the number of incoming nodes (K) is the same for all the nodes and the actual incoming connections are chosen randomly among the other $N - 1$ nodes, without repetition and avoiding self-loops. So, the variables that determine the next value of the Boolean function f_i are the values of the nodes whose outgoing arcs are connected to node i . In RBN also the Boolean functions are assigned in a random fashion. Indeed, the parameter p (called bias) defines the probability by which at each entry of the truth tables will be assigned the value 1, obviously, the $1 - p$ value determine the probability of the fraction of 0s.

Since its formulation [Kauffman, 1969b] they have distinguished themselves for the capacity of capturing relevant phenomena involved in biological cells and complex systems in general.

5.1.1 Dynamical regimes and criticality in RBN

The long-term behaviours of RBN are highly influenced by K and p values. Depending on the values of K and p it is possible to distinguish two dynamical regimes—in accordance with the theory of dynamical systems previously presented 4.3—in which the dynamics of ensembles of RBNs can be found: the *ordered* and the *chaotic*.² The dynamical regimes statistically affect the dynamics under perturbations and the length and number of attractors. In the ordered regime, RBNs are very robust against perturbations, indeed small transient perturbations die out. While in the chaotic one, they initially tend to grow, making the systems extremely sensitive to small perturbations. In the ordered regime, cycles are many more in number than in the chaotic case. Conversely, in the chaotic regime cycles are much longer.

In-depth studies [Aldana et al., 2003, Bastolla and Parisi, 1997] on the characteristics of RBNs have led to the identification of the equation that identifies the ensemble of networks in the critical regime, which separate the ordered and chaotic regimes. This equation identifies the parameters K and p that are characteristic of networks in the critical regime, networks obtained

²They are more appropriately called pseudo-chaotic since the RBN model here presented is deterministic.

with these combinations of parameters are therefore called *critical*. We can calculate the critical values of the parameter K in function of the values p (or viceversa) with the following equation, see Figure 5.3:

$$K_c = [2p_c(1 - p_c)]^{-1}$$

Networks in a critical regime have shown the most interesting behaviours. Indeed, their dynamics have properties such as robustness and adaptiveness, and for this reason—which is another proof in support of the broader conjecture that states that life exists at the edge of chaos (*criticality hypothesis* 4.3)—they have been extensively exploited to model, analyse and also reproduce in artificial context properties of living organisms, those related to cells’ dynamics in particular [Villani et al., 2011, Shmulevich et al., 2005, Huang and Ingber, 2000, Su et al., 2017].

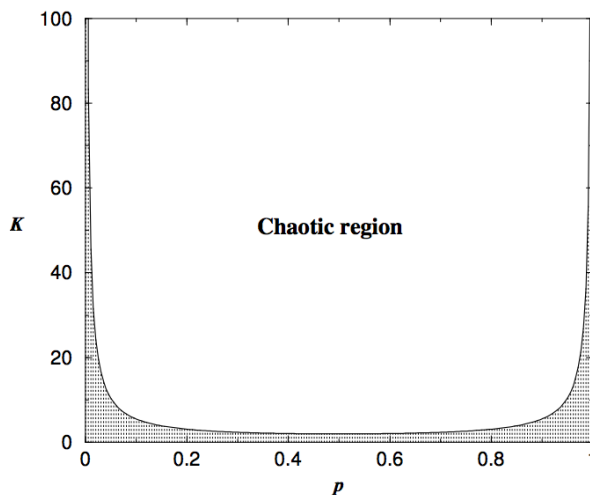


Figure 5.3: Critical line for RBN, it separates ordered and chaotic regions. Image taken from [Aldana et al., 2003].

5.1.2 Ensemble Approach

The “ensemble approach” proposed by Kauffman [Kauffman, 2004, Bornholdt and Kauffman, 2019] in its most general meaning is a suitable and powerful means to find sets—or sub-sets—of models whose properties match those of the real system of interest. The *ensembles* are defined by setting constraints on the parameters that control the creation of model’s instances, and which thus characterise the latters.

In the context genetic regulatory networks models, and therefore also in Boolean networks, this method is applied by imposing constraints in topology—which define the regulatory interactions—or in nodes functions—the genes’ rules. Examples of Boolean network ensembles are the following

(examples taken from [Kauffman, 2004]): random Boolean networks (RBN), scale free Boolean networks, “medusa” Boolean networks and others. Each ensemble shows particular statistical features—concerning the phenomenon under exam—, therefore the idea is to generate different ensembles and find out which of these correspond to or are close to the features of real cells and organisms. The ensemble(s) found in turn should predict or explain real regulatory networks behaviours or could represent a starting condition—in an evolutionary perspective—for the evolution of models towards given target characteristics.

It is precisely this twofold aspect that makes the “ensemble approach” an effective tool to investigate features of complex systems/phenomena when the knowledge about the system of interest is very limited or when the degrees of freedom and therefore the modelling possibilities are too wide to perform exhaustive investigations.

In the context of the study of cell differentiation by means of Boolean network models some important goals have been achieved through the application of this method. As an example, the generic properties of the so-called **critical ensemble** of RBNs have proven to predict features typical of the differentiation process. Very relevant is the relationship between the number of attractors in such networks and the number of cell types, as a function of the DNA content per cell and then of the nodes number. For this purpose, a recent work [Bornholdt and Kauffman, 2019] has—in the light of new genomic data that provided a better estimate of coding regions in DNA—remarked how the scaling slope between nodes and attractors number found by Kauffman [Kauffman, 1969b, Kauffman, 1993] was near to that between DNA and cell types in many phyla.

Chapter 6

Dynamical Models of Cell Differentiation

This chapter plays a central role in this dissertation because it tries to illustrate, by summarising the main works of the related literature, the conceptual steps that led to the current dynamic systems view of the biological cell and to the formulation of model of differentiation based on this view.

The idea behind this approach dates back to the 40s with the “epigenetic landscape” metaphor introduced by Waddington [Waddington, 1957]. Subsequently, with Kauffman’s works, first, and with the advent of genome-level techniques for analysing gene expression, then, the pitfalls of the classical conception of the linear causal scheme between genotype and phenotype—and hence also of the central dogma of biology—have begun to manifest themselves. Therefore, attention has shifted from the phenotypic contribution of the single gene to the dynamics of the network of interactions between regulatory genes (gene regulatory networks or GRNs). Here, the different stable asymptotic states of the GRN’s dynamics represent different cell types. Their establishment is influenced by signals or perturbations coming from neighbouring cells or—more in general—from the external environment but mainly by the constraints that the nature of interactions that compose the GRN itself imposes on its dynamics, constraints—established by the regulatory interactions—well depicted in the Waddington’s landscape [Capra and Luisi, 2014, Huang, 2012].

6.1 Cells as dynamical systems

Although some specific and recurrent biological interactions—network motifs [Alon, 2006]—present in GRNs can be well explained and understood by means of relatively simple mathematical equations, we are far from a satisfactorily understanding of the whole long-term dynamics generated by these complex networks. One possible explanation of this fact is that the

cell phenotype is not the direct consequence of the superposition of isolated genetic pathway [Huang, 2012].

Albeit abstract, a conceptual mathematical framework in which a cell is viewed as a dynamical system and its attractor states—stable equilibrium states of the GRN dynamics—underlie its observable phenotypes [Huang et al., 2009b], has been proposed since the pioneering works of Kauffman [Kauffman, 1993, Kauffman, 1969b]. This framework, that lies on the Complex Systems Science, aims to enrich the current understanding of cell dynamics and to overcome the classical linear causation scheme which links the genotype to phenotype (one gene \rightarrow one trait).

We can therefore define a network state at a given time by means of a vector state $\mathbf{S}(t) = [x_1(t), x_2(t), \dots, x_N(t)]$ where each $x_i(t)$ represents the expression level of the i -th gene which depends on the regulatory interactions between genes. So, if we represent a gene expression pattern by means of this vector state, all the possible gene expression patterns constitute the state space of the GRN and, between them, those in a **stable** equilibrium condition are attractor states, and their gene expression profiles determine the observable cell types. For any initial state $\mathbf{S}(t=0) = \mathbf{S}_0$, its *trajectory* will eventually converge to an attractor state, where the interactions forces are null.¹

6.2 Waddington’s Epigenetic Landscape

Waddington, through his “epigenetic landscape” metaphor [Waddington, 1957], had already captured with the valley abstraction the idea of basins of attraction and discrete cell fates; the marble (network state) rolls down in the landscape topology until it reaches a local minimum (the attractor state). In Figure 6.1 an adapted version of the original representation of the epigenetic landscape is reported.

We can attribute to his metaphor a formal basis and in this way explain how a network of interactions, in particular its dynamics, can give rise to a particular landscape topography. Considering that at equilibrium not all network states \mathbf{S} are equally likely, due to the interactions forces that shape the landscape, we can assign to each state a potential $V(\mathbf{S}) = -\ln p(\mathbf{S})$ where $p(\mathbf{S})$ is the probability that the network is at state \mathbf{S} when the system is at equilibrium (see supplementary material of [Huang et al., 2009b]). The function $V(\mathbf{S})$ determines the depth of the various network states in the landscape topography and the attractors states are the local minima of this function. We can think of the epigenetic landscape as the projection of the network’s state space into a plane with the valleys’ depths that represent the values of the potential function of each state.

¹Of course, as far as GRNs are concerned, not all the states might be a biologically plausible initial condition.

The model based on the attractor abstraction, of which the Waddington's landscape provides an intuitive visual representation, proved able to explain, unify and integrate various theories concerning cell dynamics in a consistent framework, free of *ad hoc* explanations.

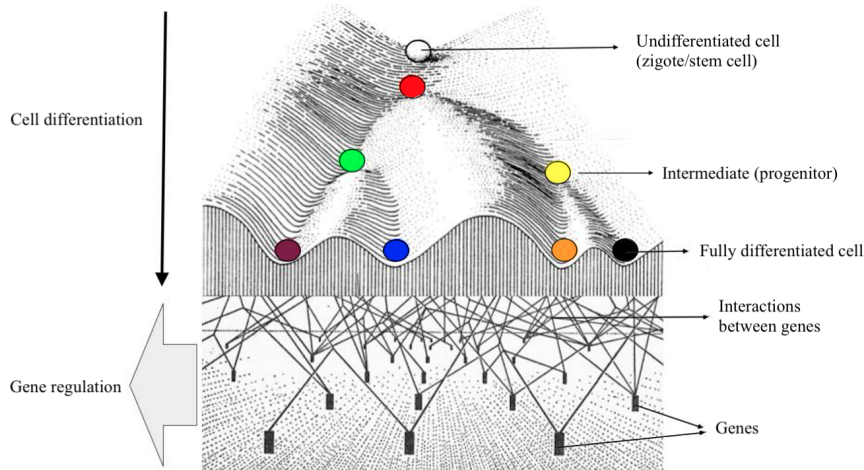


Figure 6.1: The figure has been adapted—in order to better exemplify its role in the context of the differentiation process—from the original taken from [Waddington, 1957]

Epigenetic landscape vs Epigenetics The term “epigenetics” since its introduction by Waddington have generated in the literature a lot of confusion about its meaning and what it refers to.

Without entering into the historical reasons and steps which led to ascribe the particular meanings to the term “epigenetics” and “epigenetic landscape”—which will be beyond the scope of this discussion, and they are well described in [Deans and Maggert, 2015] and in the guest essay by Patric Bateson in [Capra and Luisi, 2014]—we will report the definitions to which we adhere in order to avoid confusions and misconceptions in the remaining part of the dissertation.

Epigenetics In a manner conforming with molecular biology, with the term *epigenetics* we refer to the series of heritable mechanisms not directly derived from changes in DNA that modify the cells’ behaviour and that can persist across mitosis. Histone methylation and histone acetylation are examples of epigenetic mechanisms.

Epigenetic landscape Waddington coined the term *epigenetics* to describe “the branch of biology that studies the causal interactions between genes and their products which bring the phenotype into being” [Waddington, 1942, Deans and Maggert, 2015]. So, it referred to the processes

which a genotype undergoes during development and that ultimately determine the phenotype [Capra and Luisi, 2014]. Subsequently, the use of the expression “epigenetic landscape” established itself to refer to the original meaning Waddington attributed to “epigenetics”.

Actually, by observing the above definitions from a broader view we can state that both refer to the dynamics that take place (and the mechanisms that shape this) during developmental processes of organisms [Capra and Luisi, 2014].

6.3 Cell fates as attractors

The traditional approach to explain and understand cell regulation is based on the identification of functional signalling pathways activated by the high specificity ligand-cell surface receptor binding; this generates a cascade of signals which in turn activate specific genes for one cell fate, or more in general a cell behaviour. In the work [Huang and Ingber, 2000] the authors highlight various experimental aspects not coherent with this old paradigm because *(i)* a growth factor can induce—conversely to what is believed to be true—the activation of a very large set of genes; *(ii)* a biochemical signal can lead to different results depending on the cell state or the cell type itself; *(iii)* “non specific” mechanical stimuli can induce the same cell fates of growth factors that with high specificity bind to their receptors. These mechanisms and dynamics suggest that the cell fates are organised as attractors. In order to provide a mathematical support to this intuition and to take in account the cell fate switch produced by mechanical stimuli, they make use of a simple mathematical network model. They made a simple model of the signalling system within capillary endothelial cells including the growth factors and cell shape modulation as inputs of the model. They noticed that shape modulation in living endothelial cells produces changes within cells, related to both gene expression and signal transduction, very similar those induced by growth factors and by computer simulation of their model. These results suggested that specific molecular signals and also mechanical forces are translated into patterns of gene expression that represent attractors of the network model dynamics. Attractors are so the pre-programmed cell types or behavioural modes of the cell—growth, quiescence, differentiation, apoptosis, etc.—in which the network dynamics self-organise, relax, in the long term, also in front of different stimuli, regardless of their nature (e.g. chemical, mechanical, thermal fluctuations, etc.).

Based on previous theoretical and *in silico* results of Kauffman, Huang and colleagues in the work [Huang et al., 2005] have tried to verify, with empirical evidences, if cell types could be represented by *attractor states of the GRN*. For this purpose they stimulated *in vitro* HL60 cells by two biochemically distinct stimuli, provoking in this way initially divergent trajectories,

and observing both sets of cells converge to a macroscopically indistinguishable neutrophil stable state. Although this is only a necessary condition for the presence of an attractor state, it is particularly important because the two sets of cells follow very different gene expression trajectories before converging to the same stable state and we have to exclude the unique and common differentiation pathway hypothesis.

6.4 Cancer attractors

The “cancer attractor” concept has been firstly presented by Kauffman [Kauffman, 1971]. In brief, its conceptualisation derives from the observation that cancer cells are stable phenotypes and so they can be viewed as (aberrant) cell types that are not present in regular and healthy tissues formed during development [Huang and Kauffman, 2013]. Therefore, in the dynamical systems framework can be represented by attractor states: maybe un-evolved, unused and normally inaccessible attractors that belong to the uncharted regions of the epigenetic landscape.

Given their, apparent, uselessness in the contribution of the normal functions of organisms some speculations have been made in an evolutionary perspective on their birth and permanence as possible cellular states: it could be that they represent the price to be paid for being able to evolve new cell types [Huang, 2012].

In [Huang et al., 2009b] the authors try to contextualise the tumorigenesis within the developmental biology, avoiding the traditional vision of cancer as an aberrant product of the evolutionary process, i.e. exclusive result of genetic mutations. Explanations of cancer manifestations by means of “plausible mutations” reveal their paradoxical nature if we consider that no mutations are required to produce the various cell phenotypes generated during the development of a multicellular organism. Recalling Waddington’s metaphor, the authors propose to consider tumor types as *latent* cell types. Thus, non-genetic perturbations can facilitate cells to visit them, by placing the cell state into their basins of attractions. Remarkably this framework does not exclude genetic mutations as possible causes of tumorigenesis but relegate those to one of the possible causations of tumorigenesis, since they change the network architecture they can significantly modify the attractor landscape and facilitate the visit of cancer attractors.

6.5 A BN-based model for cell differentiation

Recently a cell differentiation model based on a noisy version of random Boolean network model (*noisy RBN*) has been proposed. This mathematical model, as we will see, is able to describe in an elegant way the most relevant features of cell differentiation. Noise plays a key role in this model;

the different stages of the differentiation process are emergent dynamical configurations deriving from the control of the intracellular noise level.

The cell differentiation model we are in this section considering has been presented in [Villani et al., 2011, Villani and Serra, 2013a, Serra et al., 2010]. This abstract model is able to describe the most relevant features of the differentiation process. We will refer to it with the adjective “abstract” because it does not refer to a specific organism or cell type. The properties of the differentiation process that which proved to be able to replicate are the following:

1. *Different degrees of differentiation*: totipotent, pluripotent, multipotent and fully differentiated cells.
2. *Stochastic differentiation*: a population of identical cells can generate different cell types, in a stochastic way.
3. *Deterministic differentiation*: activation or deactivation of specific genes or group of genes can trigger the development of a multipotent cell into a well-defined type.
4. *Limited reversibility*: a cell can come back to a previous stage under the action of appropriate signals.
5. *Induced pluripotency*: fully differentiated cells can come back to a pluripotent state by modifying the expression level of some genes.
6. *Induced change of cell type*: the expression of few transcription factors can convert one cell type into another.

This differentiation model, as we have said in brief before, is based on the so-called noisy random Boolean networks.

The kind of noise that this differentiation model takes into account is only the *intracellular* one. Indeed, it makes the following abstraction: a single cell is conceptualised as a closed system. It is generic and in principle can support different definitions of noise; however in its original formulation [Serra et al., 2010, Villani et al., 2011]—and in all our following works that are based on it—the noise type is what we will illustrate in the following.

6.5.1 Attractor transition matrix (ATM)

We investigate the asymptotic dynamics of BNs subject to noise modelled by the transient flip of a randomly chosen node which lasts for a single time step (a logic negation of node’s state). After the transient flip the BN evolves according to its usual deterministic rules until an attractor is found. This working hypothesis is made legitimate by the fact that, as reported in [Villani et al., 2011], the transitions between two different attractors very often require a small number of steps and so it is negligible the influence of the

transients: an alternative definition of “attractor” it is not necessary and we make use of the deterministic ones. This greatly simplifies the otherwise very complex dynamics that could arise, for example, by a different formulation of the concept of noise, as we will see later in 7.2. So, this noise type represents the smallest stochastic perturbation that can affect a Boolean network; even in this configuration we can observe jumps from an attractor to another one. By perturbing each node of each phase of each attractor found (one at a time), and checking in which attractor the dynamics lead we can compute the **Attractor Transition Matrix (ATM)**. This procedure is described in [Paroni et al., 2016, Villani et al., 2011, Villani and Serra, 2013a]. The ATM summarises the observed transitions between attractors and gives us an estimate of the probabilities with which such transitions can occur; a measure of the system’s robustness respect to a random flip of an arbitrary state.

6.5.2 Threshold Ergodic Set (TES)

The **Threshold Ergodic Set (TES)** is the key concept introduced in this model: indeed, cell types are modelled by TESs. A TES_θ is a set of attractors in which the dynamics of the network remains trapped, under the hypothesis that attractor transitions with probability less than threshold θ are not feasible². TESs are computed from the ATM, by iteratively removing the entries with value less than a threshold θ , which is progressively increased from 0 to 1. The TES-trees are constructed following this procedure: TES_0 represents the level 0 and each subsequent level is created if the current threshold applied to the ATM produces a different TES-landscape with respect to the previous one. In this way we capture, in a static representation, all the possible differentiation dynamics of a BN subject to noise. The static global picture of the all possible differentiation pathways that it can express and so the main characteristics of the differentiation are captured by TES-based differentiation trees, also called TES-trees. In Figure 7.4 we can find an example of a TES-tree.

The **threshold** abstraction is the principal concept introduced with this model and it plays an important role. Indeed, it is a mathematical concept strictly related with the noise level in the cell: it scales with the reciprocal of the noise level. High levels of noise (low threshold values) correspond to pluripotent cell states, where the BN trajectory can wander freely among the attractors; conversely, low levels of noise (high thresholds) induce low probabilities to jump between attractors, thus representing the case of specialised cells [Serra et al., 2010, Villani et al., 2011].

²This hypothesis is supported by the observation that cells have a finite lifetime, which enables their dynamics to explore only a portion of the possible attractor transitions.

Part III

Extensions to Current Models

Chapter 7

Extensions to BN-based differentiation models

This chapter presents a collection of works based on the powerful cell differentiation model presented in Section 6.5. In particular, we propose: automatic techniques for the design of Boolean networks capable of expressing desired differentiation dynamics, expressed through tree-like structures; evaluation of some theoretical predictions of this model (strictly related to the creation of differentiation trees) by means of stochastic simulations of BN and finally a software based on the concepts of this model for the simulation and analysis of Boolean networks.

7.1 Automatic Design of Boolean Networks for Cell Differentiation

The generic abstract properties of the model based on noisy version of RBN have been already shown to match those of the real biological phenomenon, see 6.5. A direct comparison with specific cell differentiation processes would require to design a BN (i.e. topology and node transition functions) such that its dynamics gives origin to a differentiation tree matching the properties of the real case at hand. The BN differentiation tree is characterised by the attractor set of the BN and the transitions between them, as well as their probabilities. Not surprisingly, attaining such a complex dynamics by designing a BN by hand is not possible and an approach based on brute force is definitely impractical; indeed, the number of N nodes networks with exactly k inputs per node is $(2^k)^N$. Notably, each candidate solution, i.e. a BN, is evaluated by computing its ATM, which is a highly demanding computational operation. Therefore, an automatic design method able to efficiently explore the search space is required. To the best of our knowledge, the only current method for attempting to attack this problem is a random generate and test procedure [Paroni et al., 2016], which draws BNs at random

until either an acceptable solution is found or the time limit is reached.

In this work we present an automatic design method for this purpose, based on metaheuristic algorithms [Blum and Roli, 2003]. This approach maps the BN design into an optimisation problem, where functions and topology of the BN are considered as decision variables and a measure of the matching between the BN differentiation tree generated by its ATM and a target differentiation tree is used as objective function. The objective function we defined for our algorithms is a combination of two tree distance measures: the *edit distance*, E , and the *histogram distance*, H (both distance measures have been mentioned in [Paroni et al., 2016]). The tree edit distance between two trees is the minimum cost sequence of node edit operations (node deletion, node insertion, node rename) that transforms one tree into the other¹. The histogram distance is a similarity measure between the current tree (C) and the desired tree (D), and is defined as:

$$d = \sum_{l=0}^{l^*} \sum_{k=0}^{k^*} |n_C(k, l) - n_D(k, l)| \quad (7.1)$$

where l^* denotes the maximum depth and k^* the maximum number of children nodes in both trees. The function $n_C(k, l)$ computes the number of nodes at the level l with k children in the current tree, and $n_D(k, l)$ respectively for the desired tree [Paroni et al., 2016]. In this way the histogram distance gives us a measure of the structural similarity, level by level, between the two trees; obviously, the lower the histogram distance is, the more similar two trees are. However the histogram distance might result null even if the two trees in exam are different: this may occur because this measure takes into consideration one level at a time. Several combinations of the two distances have been tested; the one leading to the best results is $F = E + (E \times H)$, which was used for the final experiments. The intuition supporting the success of this combined function is that the product between E and H initially prevails and guides the search towards regions of the landscape characterised by differentiation trees close to the target one; once this product becomes negligible, the search is then guided by E and refines the solution. A thorough landscape analysis, which would provide insights on the effectiveness of this specific combination, is subject of future work.

We devised two variants of the method, each based on a different metaheuristic algorithm; a simpler one is based on *adaptive walk*, designed mainly for test purposes and a more advanced one is implemented according to a strategy called *variable neighbourhood search*, which is capable of efficiently exploring the search space and escaping from local minima. It is important to stress that a BN whose ATM can be used to obtain a given target differentiation tree just represents one possible model for the real system to

¹<http://tree-edit-distance.dbresearch.uni-salzburg.at/>

be matched. For this reason, randomised techniques are of great help as they make it possible to explore different solutions and provide an ensemble of hypotheses. To this aim, (stochastic) metaheuristic methods are indeed particularly effective as they can be easily adapted so as to provide a wide coverage of the solutions space, e.g. by penalising already visited search space areas or by defining proper re-initialisation mechanisms that make use of some sort of memory so as to start the new search from search space areas not yet explored.

As a first step, we devise algorithms to search in the space of Boolean functions, keeping the topology of BNs constant. We consider BNs with exactly k inputs per node with random topology (without self-arcs). As experimentally shown in [Benedettini et al., 2013], this choice is not restrictive.

In the following, we illustrate the search algorithms. For more detail see [Braccini, 2016].

7.1.1 Adaptive walk algorithm

Algorithm 1 Adaptive Walk

Input: N number of nodes, K incoming degree for each node, p bias, *thresholds* thresholds list, *searchTree* desired tree, *maxIterations* number of the maximum iterations.

```

1:  $bn \leftarrow \text{GENERATERANDOMNETWORK}(N, K, p)$ 
2:  $bestNetwork \leftarrow bn$ 
3:  $tesTree \leftarrow \text{CREATETESTTREE}(bn, thresholds)$ 
4:  $distance \leftarrow \text{COMPUTEDISTANCE}(tesTree, searchTree)$ 
5:  $i \leftarrow 0$ 
6: while  $i < maxIterations$  &  $distance > 0$  do
7:    $randomFlip \leftarrow \text{GENERATEFLIP}()$ 
8:    $bn \leftarrow \text{MODIFYNETWORK}(bn, randomFlip)$ 
9:    $tesTree \leftarrow \text{CREATETESTTREE}(bn, thresholds)$ 
10:   $newDistance \leftarrow \text{COMPUTEDISTANCE}(tesTree, searchTree)$ 
11:  if  $newDistance > distance$  then
12:     $bn \leftarrow \text{MODIFYNETWORK}(bn, randomFlip)$ 
13:  else
14:     $distance \leftarrow newDistance$ 
15:     $bestNetwork \leftarrow bn$ 
16:  end if
17:   $i \leftarrow i + 1$ 
18: end while
19: return  $bestNetwork$ 

```

The AW algorithm (see Algorithm 1) performs a stochastic descent: it starts from a randomly generated BN and after the execution of each move

the resulting solution is accepted if it is not worse—w.r.t. the objective function—than the current solution. A move consists in a flip, from 0 to 1 or vice versa, of a random entry in the truth table of a randomly chosen node. So, a flip changes the genome of the gene regulatory network since it modifies the Boolean function of a node and therefore the response of a gene to certain stimuli.

Observe that this algorithm allows moves that produce solutions with values of the objective function equal to the current one, called *sideways moves*. In this way, the search is able to explore search landscape plateaus. From a biological modelling point of view, sideways moves accomplish the possibility of exploring path in the search space composed of neutral networks, i.e. different networks with the same objective function evaluation, a concept also related to genetic robustness.

The algorithm has been also optimised with a bit of memory: in order to avoid the repeated evaluation of the same network, we forbid to repeat the flip of the previous step.

The search process terminates when the objective function reaches zero (the differentiation tree found corresponds to the target one) or when the number of maximum iterations is reached.

7.1.2 VNS-like algorithm

The second algorithm we present is a metaheuristic technique inspired by *Variable Neighbourhood Search* (VNS).² This algorithm is a variant of the previously presented algorithm. AW starts with a randomly chosen network and applies an intensification strategy by making a flip to one output entry at a time. However in this way the search process, depending on the starting solution, might get trapped into local minima with no possibilities to escape or into areas of the search landscape that does not contain “good” quality solutions. For this reason we have added a diversification strategy to our algorithm. The process of diversification is implemented by increasing the number of flips if the search process does not find a solution better than the current one for a given number of steps. A better solution corresponds to a BN able to express a differentiation dynamics more similar to the desired one, i.e. with a lower value of objective function than to the one obtained by the current network. Increasing the number of random flips helps the search process to escape from local minima and it is similar to the change of neighbourhood in case of no improvements that is present in the classical VNS. As soon as a better solution is found, the number of flips is brought back to 1 and so the intensification process restarts until the objective function reaches zero or the number of maximum iterations is reached. When the number of flips is equal to 1, this algorithm behaves exactly like AW.

²See [Hansen and Mladenović, 2001, Blum and Roli, 2003] for details on this technique.

Algorithm 2 Variable Neighbourhood Search

Input: N number of nodes, K incoming degree for each node, p bias, $thresholds$ thresholds list, $searchTree$ desired tree, $maxIterations$ number of the maximum iterations, $maxNoImprovement$ number of iterations max without improvements.

```

1:  $bn \leftarrow \text{GENERATERANDOMNETWORK}(N, K, p)$ 
2:  $bestNetwork \leftarrow bn$ 
3:  $tesTree \leftarrow \text{CREATETESTTREE}(bn, thresholds)$ 
4:  $distance \leftarrow \text{COMPUTEDISTANCE}(tesTree, searchTree)$ 
5:  $noImprovement \leftarrow 0$ 
6:  $numFlip \leftarrow 1$ 
7:  $i \leftarrow 0$ 
8: while  $i < maxIterations$  &  $distance > 0$  do
9:   if  $noImprovement = maxNoImprovement$  then
10:      $noImprovement \leftarrow 0$ 
11:      $numFlip \leftarrow numFlip + 1$ 
12:     if  $numFlip > N$  then
13:       return  $bestNetwork$ 
14:     end if
15:   end if
16:    $randomFlips \leftarrow \text{GENERATEFLIPS}(numFlip)$ 
17:    $bn \leftarrow \text{MODIFYNETWORK}(bn, randomFlips)$ 
18:    $tesTree \leftarrow \text{CREATETESTTREE}(bn, thresholds)$ 
19:    $newDistance \leftarrow \text{COMPUTEDISTANCE}(tesTree, searchTree)$ 
20:   if  $newDistance > distance$  then
21:      $bn \leftarrow \text{MODIFYNETWORK}(bn, randomFlips)$ 
22:   else
23:      $distance \leftarrow newDistance$ 
24:      $bestNetwork \leftarrow bn$ 
25:     if  $newDistance = distance$  then
26:        $noImprovement \leftarrow noImprovement + 1$ 
27:     else
28:        $noImprovement \leftarrow 0$ 
29:        $numFlip \leftarrow 1$ 
30:     end if
31:   end if
32:    $i \leftarrow i + 1$ 
33: end while
34: return  $bestNetwork$ 

```

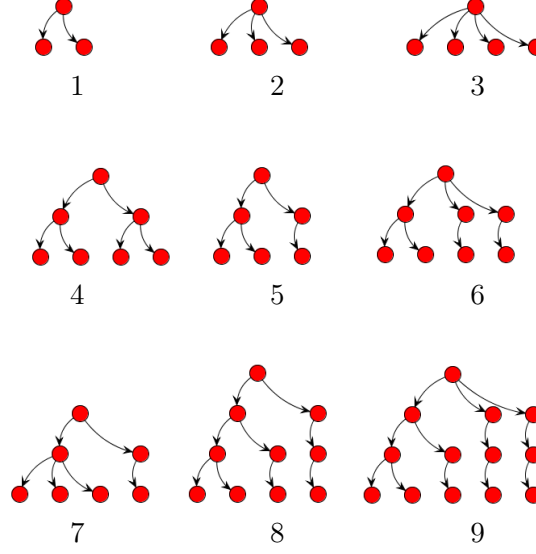


Figure 7.1: Differentiation tree structures used as target for the search process.

See Algorithm 2 for a pseudocode description of the VNS algorithm.

7.1.3 Results

We evaluated the performance of AW and VNS on target differentiation trees that were defined on the basis of common differentiation tree features, such as the hematopoietic lineage [Alberts et al., 2013]. We defined nine different tree structures, trying to capture the main features. Target trees are depicted in Figure 7.1. The threshold values at which the TESs were split have been chosen so as to be distributed in the interval $[0, 1]$ so as to capture changes in the differentiation tree corresponding to significantly different noise levels, and also to require a high probability to return to the same TESs when at a leaf of the tree. For comparison purposes, thresholds have been set to 0.2, 0.4, 0.6. The evaluation of cases with different choices for the threshold values is subject of ongoing work.

As a baseline comparison, we also run a random search that simply generates random BNs. This algorithm was allowed to generate as many networks as the maximum number of evaluations allowed to AW and VNS. For trees 1 to 7, a maximum of 10^4 evaluations has been allowed, while for trees 8 and 9, which are deeper than the previous ones, we set the maximum number of evaluations to 5×10^4 .

Experiments were run on 10 nodes BNs with $k = 2$ and with $k = 3$.

Initial RBNs with $k = 2$ were generated with Boolean function bias equal to 0.5, so as to start with BNs in the so-called critical regime [Bastolla and Parisi, 1997]. Conversely, initial RBNs with $k = 3$ were generated so that ordered, chaotic and critical regions were sampled. More precisely, we generated networks with Boolean function bias equal to 0.1, 0.5 and 0.79, respectively.

As algorithms are stochastic, 30 independent runs for each case were run and statistics were collected in case the algorithm attained at least 20 successes out of 30. Results are shown in Figures 7.2 and 7.3 by means of boxplots, which provide a visual representation of the distributions.

Trees 1, 2 and 3 are quite trivial, as they are composed of a root and some children nodes. Results on these trees are qualitatively similar: AW and VNS perform better than Random and they are equivalent in terms of number of iterations. It is interesting to observe that if initial solutions are sampled in the ordered regime, more iterations are required to design the target BNs; this fact can be explained by considering that ordered BNs have typically very few attractors, so the search has first to find networks with a sufficient number of attractors and then to modify the attractors landscape.

Trees 4 to 7 have depth 2 and are generally more difficult to be obtained, and the depth 3 of trees 8 and 9 are even more demanding. We first observe that random search is much less effective than AW and VNS, if not completely unable to find the target tree. Moreover, we note that there is no clear winner between AW and VNS. This result is a bit surprising, as we would expect VNS to be superior to AW; we conjecture that the cause of this behaviour has to be found in the structure of the search landscape, which is likely to be quite uncorrelated, which makes gradual exploration search strategies quite ineffective. An investigation of the search landscape is subject of future work.

To sum up, we observed, in general, that the success rate and the number of evaluations of the random algorithm for the most biological interesting differentiation trees are far from those reached by the other two algorithms. For this reason, we believe that additional statistical tests, which compare these two classes of algorithms, are unnecessary. As for the comparison between AW and VNS, as will be highlighted in the next section, it was not our intention to find the best algorithm, but instead, to start an analysis of possible techniques to explore the search space and find networks with common properties related to the differentiation phenomenon. For this reason, no further statistical tests were conducted to verify whether the small differences observed were due to random fluctuations.

7.1.4 Conclusion and future work

The techniques presented have shown to be superior to random search and able to robustly find BNs matching target differentiation trees with differ-

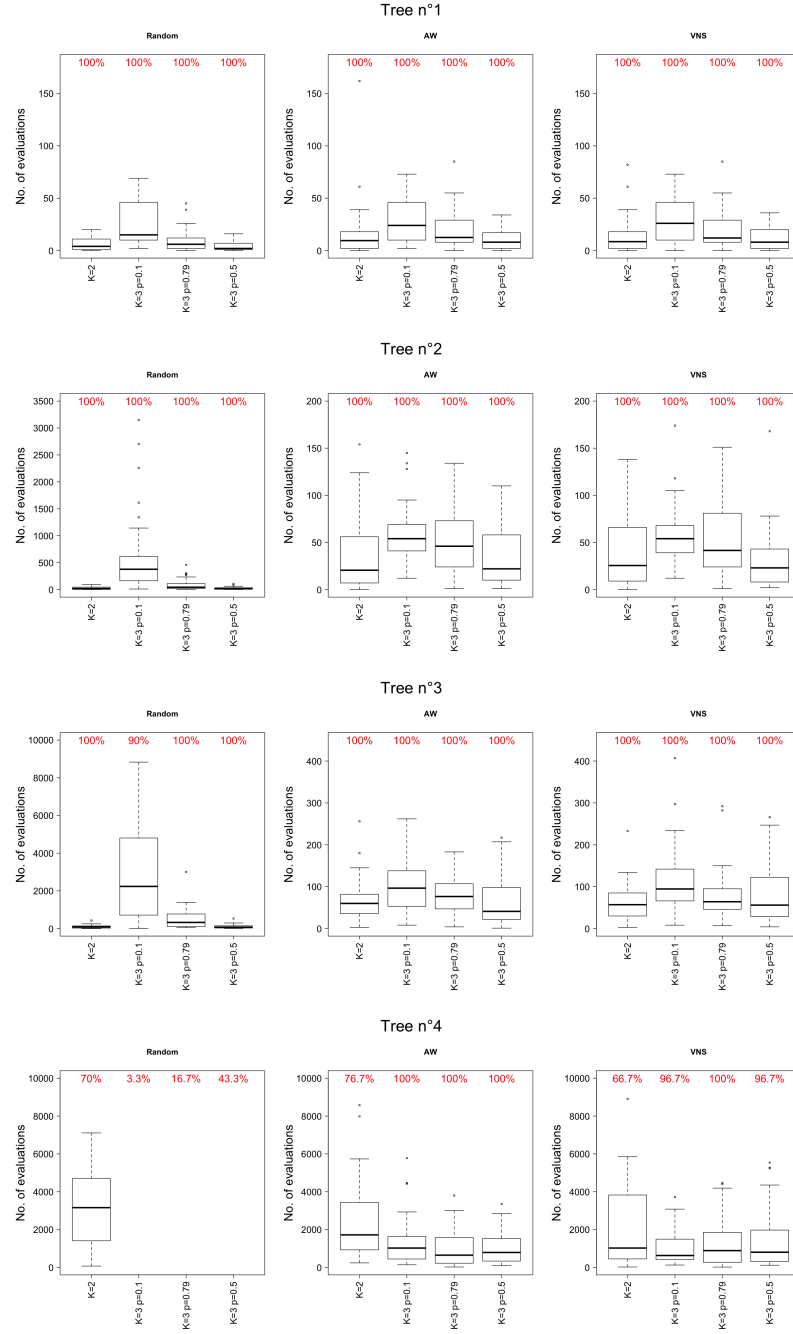


Figure 7.2: Boxplot summarising the results for differentiation trees 1 to 4. Boxplots are drawn only in those cases in which the algorithm attained at least a success ratio of 20/30. The success ratio is reported at the top of the plots.

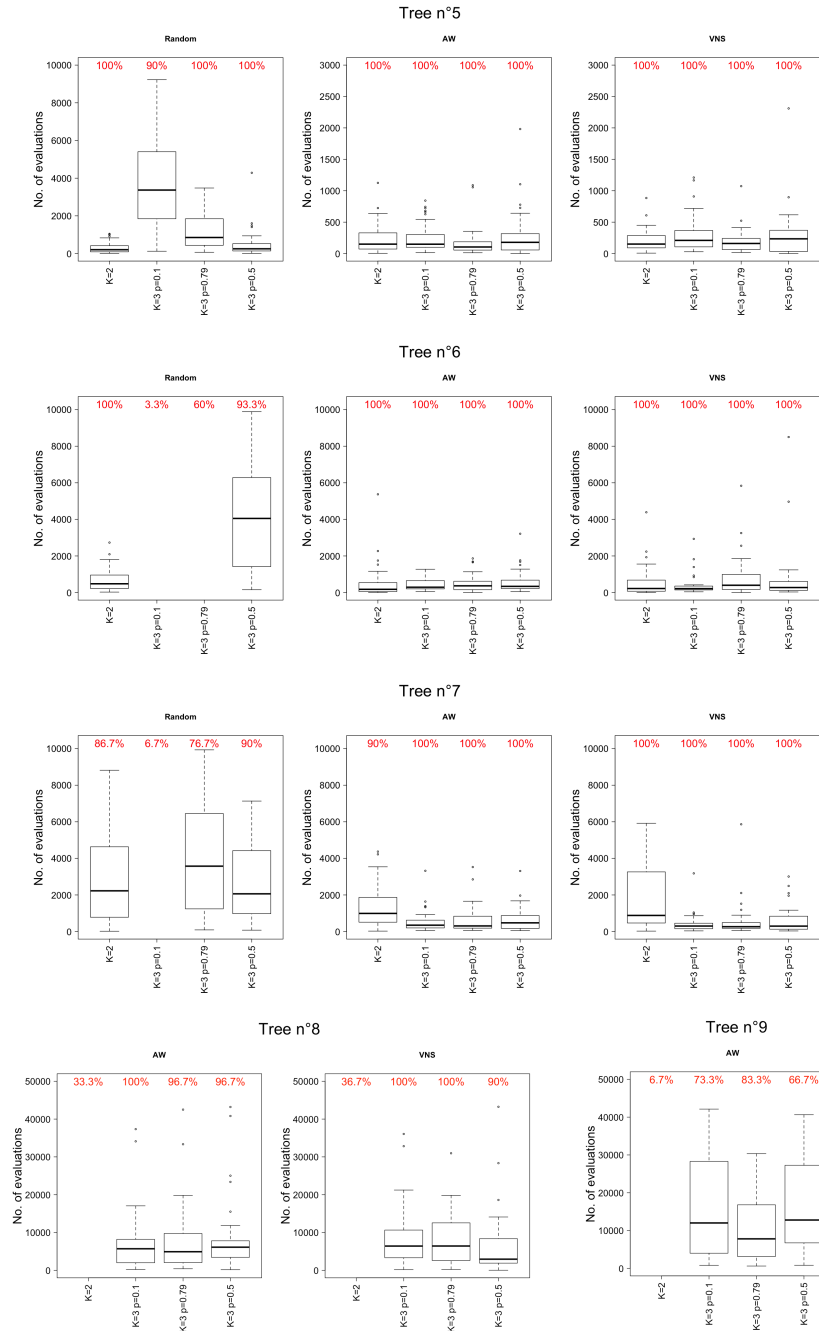


Figure 7.3: Boxplot summarising the results for differentiation trees 5 to 9. Boxplots are drawn only in those cases in which the algorithm attained at least a success ratio of 20/30. The success ratio is reported at the top of the plots.

ent characteristics. However, this work is just the first step towards the development of efficient techniques for the automatic design of BNs for cell differentiation. Since the computation of the ATM is the most costly computational operation in this process, we are trying to improve the methods so as to reduce the number of evaluations as much as possible. This may be achieved, for example, by introducing heuristics in the choice of the local moves, with the aim of performing an evaluation only for relevant moves. To tackle this problem, further metaheuristic algorithms can also be used besides AW and VNS. For example, one may want to use evolutionary computation techniques or also model-based search methods [Zlochin et al., 2004], which may be adapted so as to provide an estimation of the probability of finding a solution. Moreover, the link between differentiation tree structure and search landscape (of course depending on the objective function) has to be investigated. In fact, information on the properties of the landscape may be used to choose the solver most suited for a given tree. The approach presented in this work will also enable us to identify common features among the BNs able to produce some biological plausible differentiation trees, with the aim to find generic properties in gene regulatory networks of real organisms. In addition, following the *ensemble approach* proposed by Kauffman [Kauffman, 2004] we can generate and study different network instances and detect the properties of the ensemble that shows statistical features that match those of real cells. The techniques we propose are particularly suitable for this task as they perform a guided sampling in the BN search space and they are more efficient than random search. Finally, this approach may be extended by adding specific constraints motivated by biological plausibility, such as forcing specific activation patterns among genes.

7.2 A comparison between TESs and stochastic simulation of Boolean networks

The main contribution introduced by the model presented in the Section 6.5 is that the differentiation process is strongly correlated with the *intracellular* noise level. From the model point of view we know how the threshold is related to noise, see [Serra et al., 2010], and in addition we know that pluripotent cells have a higher intrinsic noise level than the more specialised ones [Peláez et al., 2015, Pujadas and Feinberg, 2012]. But the threshold and above all its variation mechanism introduced in the model (with which we model the differentiation process) are *externally controlled*. In fact the threshold represents an abstraction of the mechanisms implemented by the real cell to control noise. The identification of autogenous mechanisms, somehow bound to cell's dynamics, through which achieve a threshold self-regulation is subject of ongoing work. As first step to identify the biological mechanisms that affect noise level, and in turn the threshold, we can take

in exam a system with different types of noise and noise levels and we can verify if the system is able to reproduce the TES phenomenology. In fact, the approach to cell differentiation previously presented might not capture the real asymptotic configurations of real cells if the cellular system is subject to a noise implemented in a different way with respect to the original model. For example, a real cell dynamics might quickly diverge from the TES model's prevision if its dynamics is such that:

- more than one noise events can occur simultaneously in an asymptotic state;
- noise events occur in its transients.

In addition, the TES-based differentiation trees are constructed following a specific process of threshold variation on the ATM. This process allows us to observe all the differentiation pathways the GRN model is capable of expressing, under a particular noise setting.

To verify to what extent can the TES model predict the entire spectrum of scenarios produced by the dynamics of a system subject to intrinsic noise, we perform time evolutions of Boolean networks subject to different noise levels and we compare these two approaches. Noise levels are represented by distinct frequencies of random perturbations. In such a way, we have the means for counting—for each noise level—the number of differences between the outcomes obtained with the TES model and the stochastic simulations. In the following we call a *story* a single time evolution of a BN subject to random perturbations. Considering that we are interested in the asymptotic behaviour of the BN dynamics we count the jumps between attractors obtained in each story and we compare them with each level of the TES-based differentiation tree, computed using the TES-model approach on the same BN. We call an *incompatibility* a jump between attractors that would not be allowed given the TES-landscape of a tree's level.

7.2.1 Experimental setting

The Boolean networks used in the experiments have $n = 100$ nodes and $k = 2$ distinct inputs per node assigned randomly (self-loops are not allowed). Boolean functions have been set by assigning a 1 in the node truth table so as to attain exactly a frequency of 0.5 across all the truth tables (for $k = 2$, this corresponds to the critical value [Bastolla and Parisi, 1997]). The rationale behind this choice is that in preliminary results, by setting the bias for each Boolean function, in some instances the average overall bias calculated on all nodes could have a non-negligible standard deviation from the desired mean value. Because we want to estimate the differences between the model and the stochastic simulations, we did not want the results to be affected by variance in network dynamic regime. So we use an exact bias, following

this procedure for generating networks: we generate a vector of length equal to the sum of the number of Boolean functions' entries of all nodes in the network ($2^k * n$), we assign half the values to 1 and half to 0 and we use a random permutation of this vector to define the Boolean functions.

The BN is subject to a synchronous dynamics, i.e. all nodes update their state in parallel and functions are applied deterministically. Given that the typical time needed to transcribe a gene is equal to 25 - 50 sec in yeasts and 2 - 3 min in mammals (see reference BNID 111611 [Milo et al., 2009]); we assume 1 minute as a plausible mean value for a BN's synchronous step of update. In addition, analysing the cell's average life span in humans (see reference BNID 101940 [Milo et al., 2009]) we set to 5×10^4 the number of steps for a BN run, in order to model an upper bound of plausible mean cell lifetimes (approximately one month). The only stochastic component resides in the noise, which has been simulated as a temporary flip of the value of a node applied with probability ν ; hence, at each step of the temporal evolution of the network, νn nodes are flipped **on average**. We ran experiments with ν so as to have on average one flip every τ steps, with $\tau \in \{1, 5, 10, 15, 20, 50, 100, 200, 500, 10^3, 5 \times 10^3, 10^4, 2 \times 10^4, 5 \times 10^4\}$. In the following, we will denote the corresponding noise probabilities as ν_τ . Note that the higher τ , the lower the probability ν applied to each node. This noise mechanism emulates possible temporary fluctuations in the expression level of genes and may occur both during stationary phases (i.e. along attractors of the BN) and transients. We run experiments with 30 random BNs; for each of them we compute the ATM and then the TES-tree, following the procedures mentioned in Section 6.5. A typical TES-tree is depicted in Figure 7.4. The time evolution of each BN was also simulated 100 times (100 *stories*), each one of them starting from a random initial state. We collected the trajectories of the BNs and computed statistics on the compatibility between the stories and the TES-tree, besides other ancillary statistics on the overall dynamics of the BNs.

7.2.2 Results

In this section we provide the results obtained. The comparison between TES-trees and simulations with stochastic noise is mainly based on counting the transitions between attractors that are observed in the stochastic simulation but that are not allowed by the ATM, given a probability threshold θ . That is, the analysis of what we have called *incompatibilities* between the two approaches for modelling cell differentiation. For each value of ν_τ , we counted the incompatibilities observed in all the 100 stories w.r.t. the lowest non-zero value of θ (level 1 of the TES-tree) and the highest one, where all TESs are single attractors (level n of the TES-tree). These two particular levels are taken as representative elements able to summarise the trend of incompatibilities since level 1 represents the first TES with not trivial con-

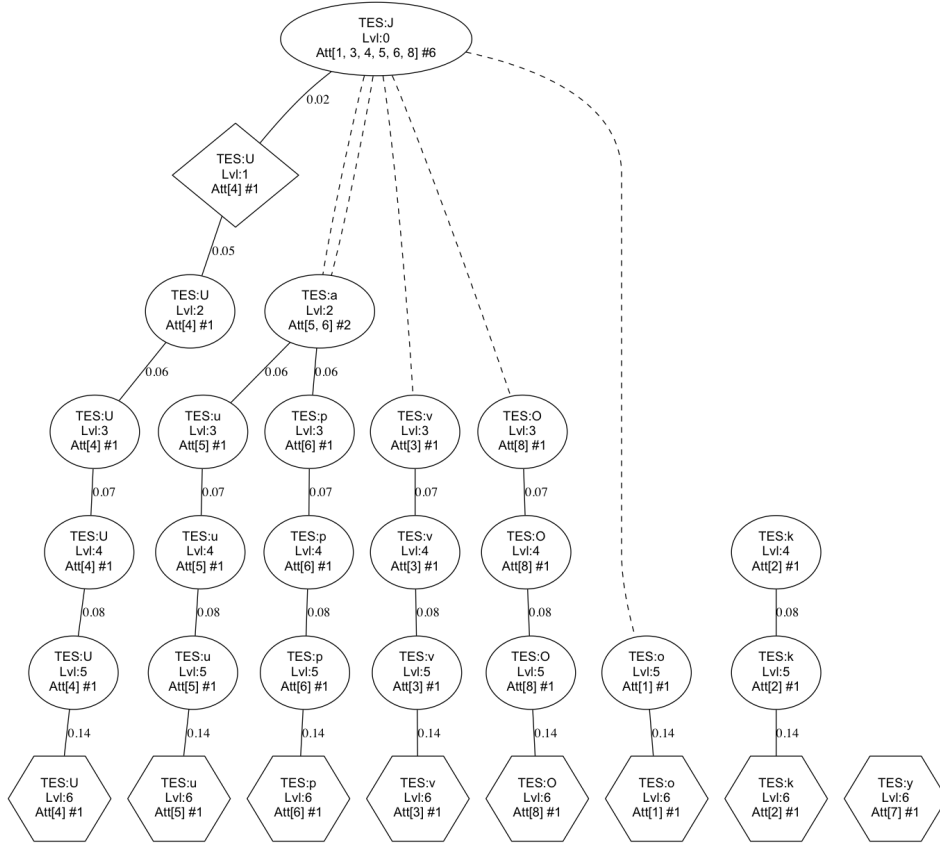


Figure 7.4: An example of a TES-tree. Levels are numbered from 0, the topmost, to n , the lowermost; $n = 6$ in this example. TES of level 1 has a diamond shape whereas TESs of level n have an hexagonal one. Labels on the edges indicate the minimum threshold value at which any TESs of the previous level splits or reduces. Continuous lines denote paths along the differentiation tree that can be followed by increasing the threshold at minimum steps (these values are directly obtained by the ATM). Dashed lines denote the paths that can instead be followed if the threshold was increased by larger steps.

straints and level n is the most constrained one. Results are summarised in Figures 7.5, 7.6 and 7.7. In these figures the boxplots graphically represent distributions of the median values of the overall incompatibilities (computed on all 30 BNs) with respect to a particular noise level; different noise levels are represented by distinct colours. For each noise level two boxplots are plotted, one for the incompatibilities with respect to the level 1 and one for the level n .

As expected, the higher ν_τ (corresponding to low values of τ), the higher the number of these incompatibilities. Moreover, this increases with θ ; which corresponds to the increase of the TES-tree's depth. Despite the discrepancy which is apparent at high noise levels, we observe that already for medium noise levels, i.e. not higher than ν_{200} , the incompatibilities are limited and tend to be negligible towards low noise levels.

As previously stated, we could observe marked differences between model and simulations if the actual noise presents in the stories is different from that hypothesised by the model. Hence, we analyse the dynamics of the stochastic simulations and we count the number of noise events occurred during transients and the multiple flips in attractors. With multiple flips we mean the occurrence of more than one node value change at a time. Situations both not covered in the model and which could represent the main causes of divergence between the two approaches. In Figures 7.8, 7.9, 7.10 and 7.11 each distribution summarises the median values of the property in exam; the median value for each BN computed across the 100 stories of a particular noise level. Hence, we have one boxplot for each distribution of medians. These statistics show that noise events in transients and multiple flips decrease in an exponential way as noise decreases. This trend is more evident in figures 7.9 and 7.11, which have logarithmic scales. We can note that under noise level ν_{100} the number of multiple flips and noise during transients become negligible with respect to the number of steps considered in the stories (i.e. 5×10^4). We must remark that although the flip of a gene is the smallest stochastic perturbation that can affect a Boolean network it biologically reproduces a fairly intense event, much stronger than molecular fluctuations. Hence, the noise level ν_{200} (250 noise events on average in a story) identified as the convergence point between the two approaches could even be a too high noise level for a real cell's life span. This observation contextualises the results obtained in a biological framework and it highlights the relevant noise levels in which a real cell can operate.

The results obtained support the statement that there exists a significant noise level under which the two models are in agreement. Therefore, *(i)* under this threshold they can be both used to model differentiation phenomena—and their observations can be combined—and *(ii)* the new dynamic simulations may add interesting pieces of information on the heterogeneities of the possible individual configurations.

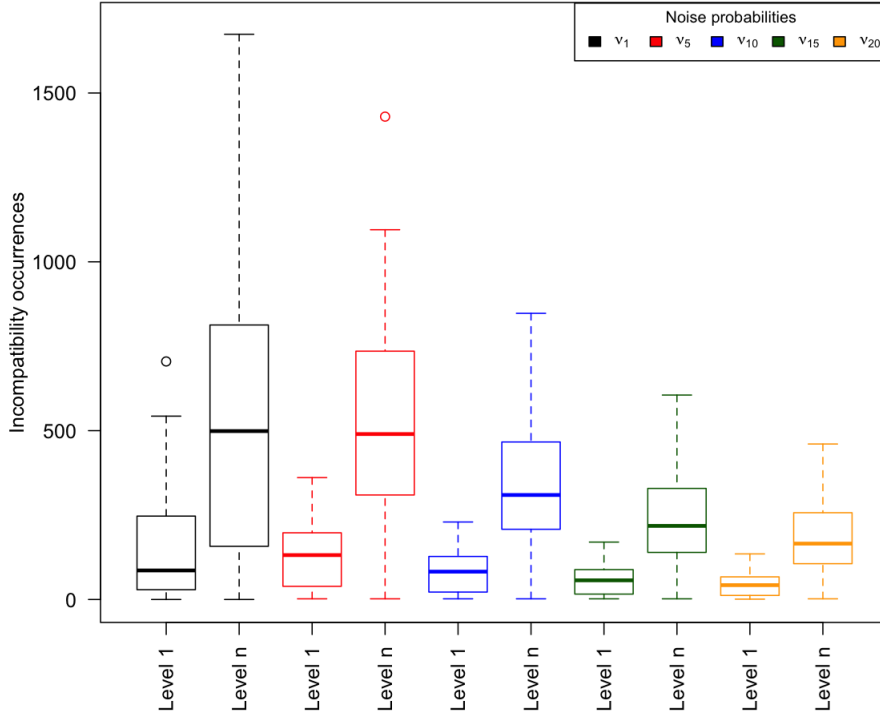


Figure 7.5: Distribution of the median values of the *incompatibilities* between the level 1 and level n of the TES-trees and stochastic simulations with the probabilities to flip a node ν so as to have on average one flip every 1, 5, 10, 15, 20 steps. Noise probabilities ν_τ expressed with different colours.

7.2.3 Conclusion

In this work we have compared two approaches for modelling cell differentiation, both based on random Boolean networks subject to noise. One approach is represented by the well-known model based on TES concept, the other is grounded in time evolutions of BNs subject to different noise levels. The analysis of the emerging differences between these two approaches suggests that there is a specific noise level under which the two models produce similar results. This result has important implications because it shows that both approaches can be used to model cell differentiation and in addition their outcomes can be, at least in part, complementary. Indeed, the new approach could be used to determine the distribution of the extra-cellular noise, due to the intra-cellular events. Moreover this work produced, on the one hand, another proof of robustness of the TES-based differentiation model and, on the other, since the stochastic simulations of BN require less

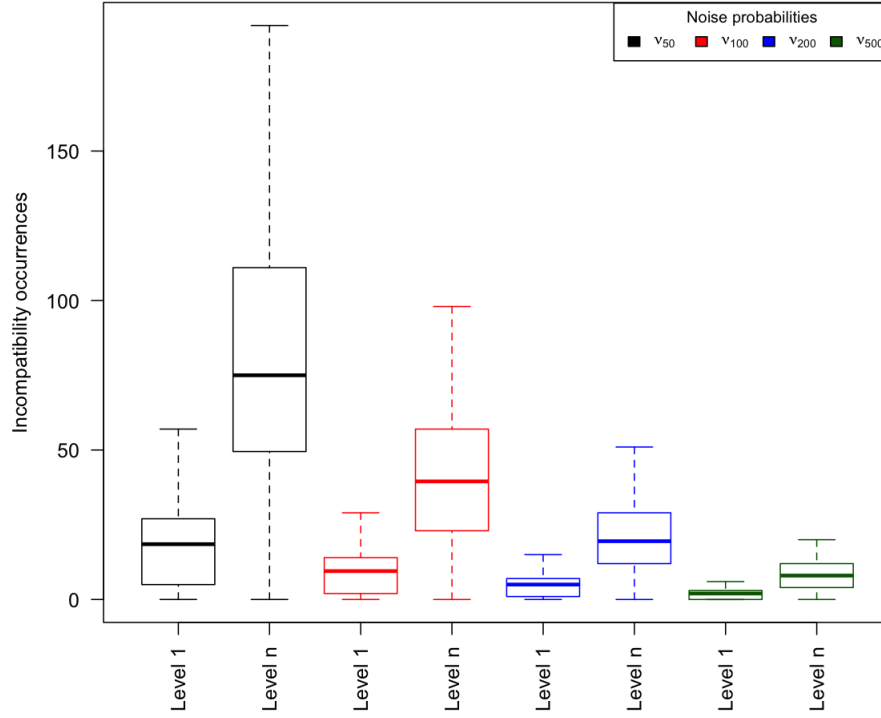


Figure 7.6: Distribution of the median values of the *incompatibilities* between the level 1 and level n of the TES-trees and stochastic simulations with the probabilities to flip a node ν so as to have on average one flip every 50, 100, 200, 500 steps. Noise probabilities ν_τ expressed with different colours.

computational cost than the TES model they can be used as an alternative and exploitable approach to conceive more performing automatic procedure for generating biologically plausible cell differentiation model based on BNs [Braccini et al., 2017, Benedettini et al., 2014].

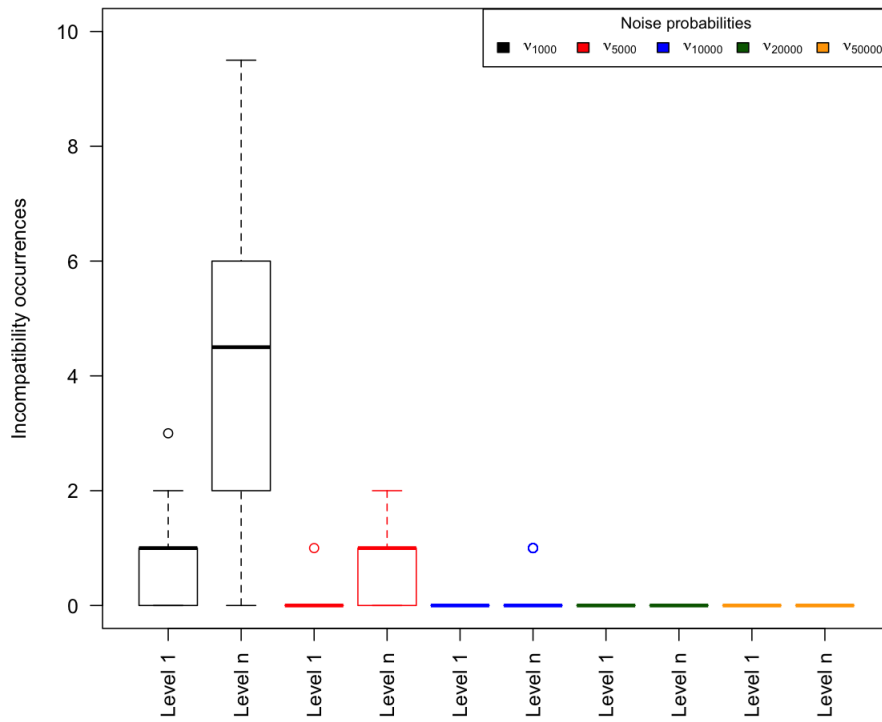


Figure 7.7: Distribution of the median values of the *incompatibilities* between the level 1 and level n of the TES-trees and stochastic simulations with the probabilities to flip a node ν so as to have on average one flip every 1000, 5000, 10000, 20000, 50000 steps. Noise probabilities ν_τ expressed with different colours.

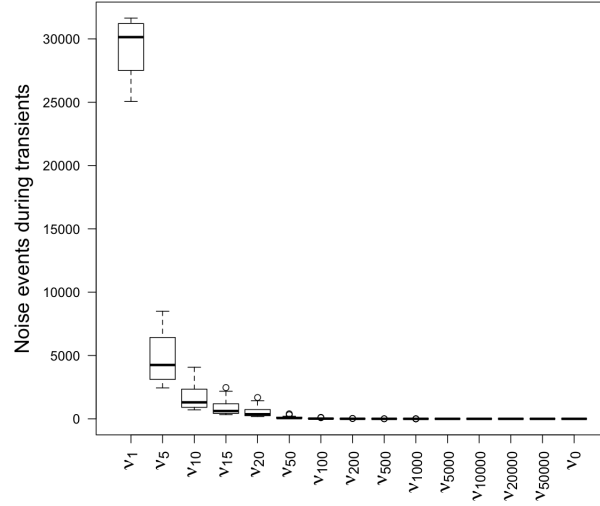


Figure 7.8: Distribution of the median values of the number of noise events occurred during transients in stochastic simulations (*stories*), for different noise levels. Noise levels expressed by the ν_τ values in the x axis.

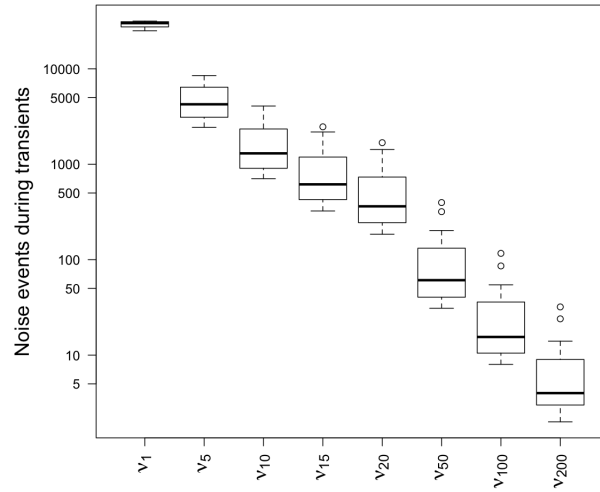


Figure 7.9: Detail of Figure 7.8 on logarithmic scale.

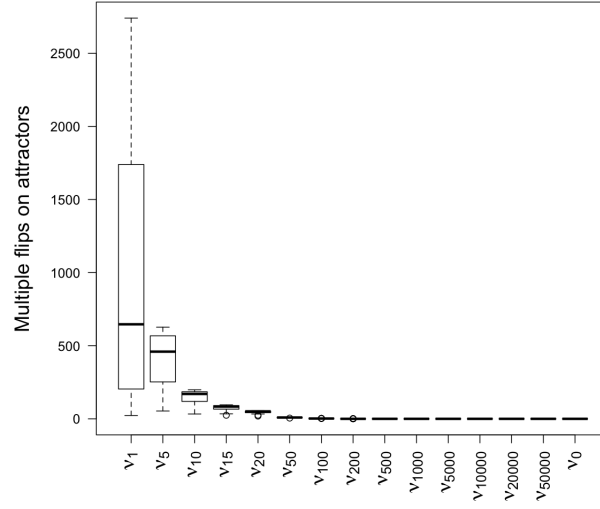


Figure 7.10: Distribution of the median values of the number of multiple flips occurred in the attractors in stochastic simulations (*stories*), for different noise levels. Noise levels expressed by the ν_τ values in the x axis.

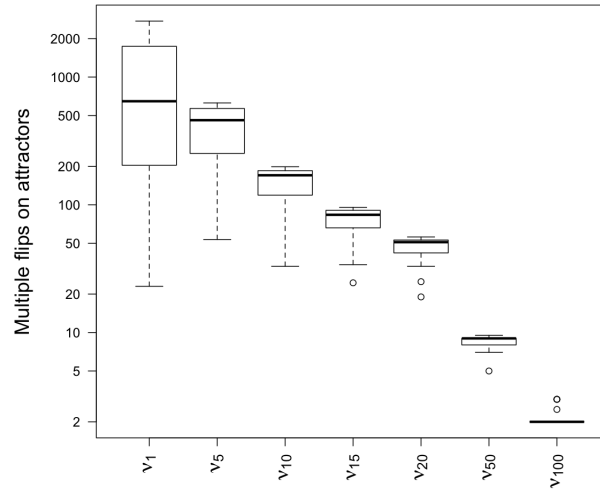


Figure 7.11: Detail of Figure 7.10 on logarithmic scale.

7.3 A software library: *diffeRenTES*

So far the dynamical model for cell differentiation introduced in [Serra et al., 2010, Villani et al., 2011, Villani and Serra, 2013b] and summarised in Section 6.5 has been implemented only in CABeRNET [Paroni et al., 2016], a Cytoscape application. But since more and more biologists—and in general scientists without a computer science background—are approaching the world of biological modeling, it is necessary to provide them with software and tools that are simple to use but at the same time effective for the problems tackled. Since R is a statistical software very used by computational biologists and with a very good package for Boolean networks simulation and analysis, namely *BoolNet*, we thought that it was important to provide to the R community a package for modeling cell differentiation with BN.

Hence, *diffeRenTES* (from French various, different, separate) is a package written in R for computing, starting from a Boolean network, the main abstractions on which the dynamical model for cell differentiation is based on. In particular, the package is able to compute the ATM (Attractor Transition Matrix) structure and tree-like structures based on the TES (Threshold Ergodic Sets) concept that describe the process of differentiation.

The main functions of the library are here reported, accompanied by their signature:

- `getATM <- function(net, syncAttractors, MAX_STEPS_TO_FIND_ATTRACTORS = 1000)`
- `getTESs <- function(ATM)`
- `getDifferentiationTreeAsDOTString <- function(TESs)`

The *getATM* function is the starting point of all computation that it is possible to carry out with this library. For computational reasons, it relies on the previously computed set of attractors with a synchronous updating scheme (*syncAttractors* parameter) of a specified Boolean network (*net* parameter) by means of the well-known *BoolNet* [Müssel et al., 2010] library, also available in R. In addition it defines a formal parameter *MAX_STEPS_TO_FIND_ATTRACTORS* with a default value of 1000 with which it specifies the limit number of steps before interrupting the search of the reachable attractor after a perturbation event.

The *getTESs* function produce a static global picture of the differentiation process, summarising all the possible outcomes of the BN dynamics under a particular noise setting, the same noise setting used for computing the ATM (*ATM* parameter). The TESs returned are constructed following a specific process of threshold variation on the ATM: the cell differentiation dynamics is expressed as the time evolution of BNs subject to progressively decreasing noise. This process allows us to observe all the differentiation pathways the GRN model is capable of expressing.

The `getDifferentiationTreeAsDOTString` method produces a string—in DOT format³—that represents the TES-based differentiation tree.

Here, a simple example of usage of the package is presented: the R listing makes explicit use of the BoolNet library for the Boolean network random generation and for the attractors' computation.

```
net          <- BoolNet::generateRandomNKNetwork(10, 2)
attractors   <- BoolNet::getAttractors(net)
atm          <- getATM(net, attractors)
tes          <- getTESs(ATM)
s            <- getDifferentiationTreeAsDOTString(tes)

# Writing a file containing the DOT representation of
# the TES-based differentiation tree
write(s, "example.gv")
```

At the time of this writing, the library is released under GPL-3 license and it is available on GitHub at the following link <https://github.com/mbraccini/diffeRenTES>. In the future, we would like to publish the package in the CRAN repository <https://cran.r-project.org/> in order to make available to a greater number of scientists. To this extent, the package has been developed following the state-of-the-art software engineering techniques and technologies, like continuous integration software (*Travis CI* in this case) and software for teamwork (*Git*).

³<https://www.graphviz.org/doc/info/lang.html>

Chapter 8

A network motif in BN: Self-loops

8.1 The impact of self-loops in RBN dynamics: A simulation analysis

Since completely random networks are unrealistic, some work has been done to extend the original model with structural and functional properties observed in biological networks. Among recurring motifs identified by experimental studies, auto-regulation seems to play a significant role in gene regulatory networks. We, therefore, have introduced a model of auto-regulatory mechanisms by introducing self-loops in RBNs. By means of *in silico* experiments we analysed the impact of self-loops in the RBNs asymptotic behaviour. In particular, we performed different simulation experiments where a RBN created completely random is incrementally modified introducing one self-loop at a time until every node has one. Different configurations, in terms of network topology and functions, are evaluated.

8.1.1 Motivation and goal

When generic properties are sought, the typical approach consists in studying ensembles of Boolean networks generated according to a given, biologically plausible, model, such as the one proposed by Kauffman [Kauffman, 1993]. In this model a Random Boolean Network (RBN) is initialised completely random both in the topology and in the functions, possibly defining the number of inputs each node has. Variants of this model have also been considered, for example by restricting the set of Boolean functions to canalising ones, or by imposing a scale-free topology. These variants are inspired by biological plausibility and are often suggested by the identification of crucial properties and mechanisms observed in GRNs reconstructed from biological data.

Extensions investigated, for instance, the role of noise in the stability of the asymptotic states [Villani et al., 2011, Hoffmann et al., 2008], since it plays a crucial role in cellular regulatory networks [McAdams and Arkin, 1997]. Other works [Harris et al., 2002, Kauffman et al., 2003] discuss whether Boolean rules, if selected randomly among all the possible Boolean functions of k inputs, are an acceptable approximation to model gene regulation. Their conclusion states that canalising rules reproduce experimental observation more accurately. Finally, the distribution of GRNs is analysed, observing that they mainly exhibit a scale-free topology [Albert, 2005], and simulation experiments with scale-free RBNs are performed, suggesting that nor pure random neither scale-free are likely the best approximation for GRN topology and that further studies are worthy [Aldana, 2003, Serra et al., 2004a, Serra et al., 2008].

The contribution here presented has the long term goal of identifying basic mechanisms and common motifs of GRNs underlying fundamental cellular processes [Yeger-Lotem et al., 2004, Shen-Orr et al., 2002] that can be modelled as structural and functional elementary bricks in BNs, thus making it possible to study generic properties of cell dynamics by means of ensembles of more realistic BNs models. We believe that analysing common motifs of GRNs, and identifying the best solution for modelling them as structural and functional elementary bricks in RBNs, would support a more aware analysis and understanding of the emergent dynamics obtained with the simulation [Ahnert and Fink, 2016, Yeger-Lotem et al., 2004]. Furthermore, this repertoire of bricks may be used inside algorithms for the automatic generation of BNs endowed with specific dynamical properties [Benedettini et al., 2014, Braccini et al., 2017]. Such networks may also be exploited for designing and controlling the behaviour of artificial entities [Francesca et al., 2014]. Our long term goal is to build a catalogue of bricks—similarly from the BioBricksTM idea of Synthetic Biology [Shetty et al., 2008]—whose function and role inside a GRN is known. The expected impact of this bricks catalogue is twofold. On one side the analysis of GRNs dynamics via simulation will provide an in depth clue on the link between the function of the parts and the emergent behaviour observed. On the other side, the engineering and design of RBNs with specific behaviours will then be possible by composing known bricks. This work represent a first step towards this challenging result.

As a first step, we have analysed the role and the impact of self-loops, which abounds in biological genetic networks, in RBNs dynamics. Indeed, within a GRN, a self-loop models the property of a gene producing some chemical substances that contribute at the regulation of its own gene. In particular we focus here on positive self-loops, whose effect in Boolean models is to maintain the activation state of the gene. The network motif under study here is *self-loop*, also known as auto-regulation mechanism, *i.e.*, the gene regulation motif where a transcription factor regulates the transcription

of its own gene. Self-loops abound in biological genetic networks [Hermesen et al., 2010]. In this work we focus on positive self-loops responsible for the up-regulation of their own genes. This mechanism is particularly evident in the differentiation process, where cells, from a stem state, choose a fate towards specific specialised cells. For example, from the *Drosophila* GRN shown in Fig. 8 of [Montagna et al., 2015], all the four main genes responsible for the patterning of gap genes expression during embryo development are involved in autocatalytic reactions. To the best of our knowledge, the impact of self-loops has only preliminarily been studied in RBNs. A first work is presented in [Pinho et al., 2014], where the relation between the sign of the regulation (positive or negative) and the robustness of the network is investigated.

8.1.2 Methods

The impact of the introduction of self-loops into a RBN has been studied through simulation. In particular the goal was to observe how the asymptotic behaviour of the network, namely the number of attractors and their stability, changes as a function of the fraction of self-loops added. As in the RBN model introduced by Kauffman [Kauffman, 1969b], we suppose that one node in the RBN corresponds to one gene in the GRN. For simulation purposes, we modified a randomly generated Boolean network in different ways; we have:

AUGM-RND: added a self-loop and extended the truth table randomly (with the same bias used for generating the original RBN);

AUGM-OR: added a self-loop and changed the node Boolean function into an OR between the node value and the previous function;

CONST-RND: removed an incoming link and replaced the input with a self-loop, without changing the node Boolean function;

CONST-OR: removed an incoming link and replaced the input with a self-loop, and changed the node Boolean function into an OR.

In Figure 8.1 we elucidate how a single network node is modified to obtain the experiments configurations previously mentioned. We thus explore the role of self-loops in both the cases of maintaining a random Boolean function and of adopting a canalising function (OR). The choice to explore canalising functions is motivated by the focus on self-loops with self-activating effect. Indeed, according to the role they have in biological networks, a positive

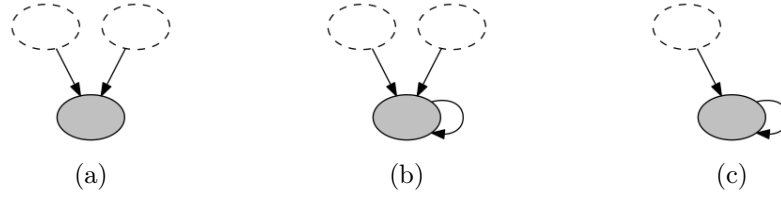


Figure 8.1: a: Typical generic node generated following the original RBN model with $k = 2$, self-loops are not allowed. b: Example of a node modified for experiments with a self-loop added. c: Example of a node modified for experiments with an incoming link removal and self-loop addition. Outcoming arcs are not drawn since both the RBN model and our models do not impose constraints on them.

self-loop should model the property of a gene producing some chemical substances that contribute maintaining the activation state of that gene. This means, within a RBN, that the function should keep the node value. It is worth mentioning that, since within a RBN we do not associate 0 with gene off and 1 with gene on, but we are interested in maintenance and transition of states, using an OR function or an AND function is conceptually the same.

For each of these experiments, self-loops are introduced incrementally to the original RBN. In this way we have a fraction of self-loops varying from 0 to 1 (no node has a self-loop – all nodes have a self-loop) and we are able to observe how the behaviour of the network is modified step-by-step.

In all experiments, each RBN is simulated following a synchronous dynamics update scheme—*i.e.*, nodes update their states at the same instant—and with deterministic functions. Since the state space is finite, the BN after a transient eventually reaches a *fixed point* or a *cyclic attractor*; these are the only achievable asymptotic states in this setting.

Statistics are taken across 50 different RBNs with $n = 20$ nodes. Initial RBNs are created with $k = 2$ and function bias p equal to 0.5. The value of the k parameter is chosen for its biological plausibility [Kauffman, 1969b, Kauffman, 1969a, Serra et al., 2004b]. The ensemble of RBNs having these bias and k values are in critical dynamic regime [Bastolla and Parisi, 1997]. We sampled from this networks ensemble because, statistically, they exhibit robustness and adaptiveness similar to real genetic regulatory networks [Villani et al., 2011, Shmulevich et al., 2005]. In the experiments, we explored all the possible initial states (2^{20}) of the RBN, to obtain the whole attractors landscape. Since we did not want the results of comparisons between models to be affected by the variance of the network dynamic regime, we set an exact bias to initial RBNs (before any modification). To do this, we computed a random permutation of a vector with length equal to the sum of all Boolean functions entries ($2^k * n$) with half of the values to 1 and the remaining to 0; we take a portion of this vector to populate the truth table

of a node. To estimate the dynamic robustness of a network we introduced noise, modelled by a random flip of a randomly chosen node (logic negation of a node state). We have flipped each node of each state of each attractor in order to compute the ATM 6.5.1. In particular, to compute our statistics measuring the network robustness, we examine the main diagonal of the ATM: each diagonal entry give us the estimate probability of returning in the same attractor as a result of a random node perturbation.

8.1.3 Results

Results reveal that self-loops massively affect the number of attractors and their robustness. Figure 8.2, 8.3, 8.4 and 8.5 show how the average number of attractors and the probability of returning to an attractor vary as a function of the fraction of self-loops. In particular, on the left side of Figure 8.2, 8.3, 8.4 and 8.5, each point corresponds to the average number of attractors obtained across the 50 networks with a particular fraction of self-loops. On the right side, we have the robustness trend as a function of the fraction of self-loops; each boxplot represents the distribution of the ATM main diagonal values from all 50 BNs.

Generally speaking, we can observe that the number of attractors is higher in the networks with self-loops.

It grows quasi-exponentially, thus the effect observed is gradually more evident with increasing number of self-loops. In particular under around 30% of self-loops, the number of attractors slowly grows, not impacting too much the network dynamics, while afterwards it sharply increases, strongly reconfiguring the attractor landscape. At the same time, attractors' robustness tends to be smaller than in classical RBNs. This result is quite intuitive: since the number of attractors is significantly higher, the size of the basins of attraction should on average be smaller, thus making less likely to return to the same attractor after a flip, while easier to move in an other attractor. Moreover, the impact is even more striking when Boolean functions are changed into an OR between the previous function and the node value involved in the self-loop.

Even though we have found that the results are qualitatively the same for all the four variants we considered for introducing self-loops (as discussed in Section 8.1.2), in the following we detail the main differences we observed.

AUGM-RND: the number of attractors varies quasi-exponentially until around 70 attractors for networks with a self-loop in each node. Conversely the median value of the returning probability decreases from around 0.6 to 0.3. However the distribution of the ATM main diagonal values is widely distributed between the extreme values of the range;

AUGM-OR: the peculiar characteristic of this experiment is that the num-

AUGM-RND

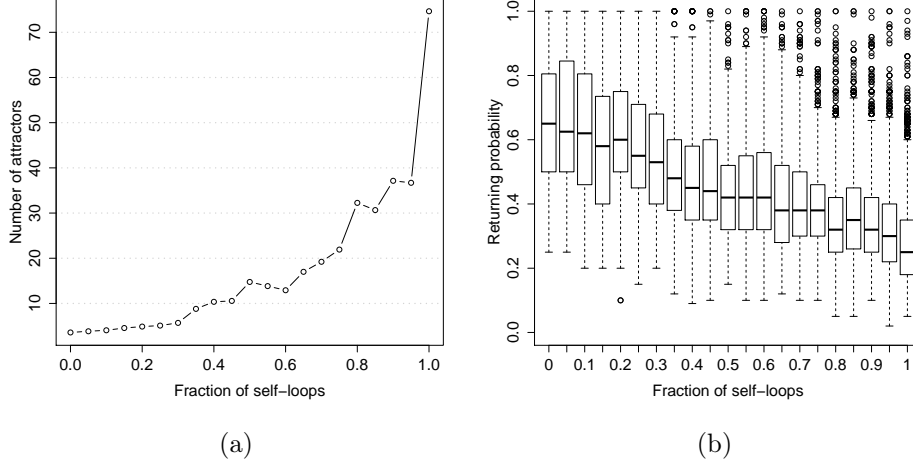


Figure 8.2: Average number of attractors (*i.e.*, both cycles and fixed points) as a function of the fraction of self-loops added in RBNs originally with $k = 2$ (8.2a). Distribution of the probabilities of returning to an attractor after one node flip (8.2b).

ber of attractors raises exponentially until around 2000 attractors. We used a logarithmic scale to zoom the plot to just a few nodes with self-loops; we motivate this significant difference in the attractors number, with respect to the previous configuration, noting that the OR function increases the probability that one node is in the 1 state, and it assures that it remains at that value;

CONST-RND: the number of attractors varies approximately exponentially until around 400 attractors, *i.e.*, more than the max number of attractors we have in experiments where we add a self-loop;

CONST-OR: the number of attractors varies exponentially until around 600 attractors. Peculiar in this setting is the median value of the returning probability graph where the probabilities of returning to an attractor, after one node flip, decreases until around 0.1, which is the lowest value we have in all the settings we considered.

8.1.4 RBNs with OR functions

A question may be asked as to what extent the observed results depend on OR functions rather than self-loops. To address this question we first observe

AUGM-OR

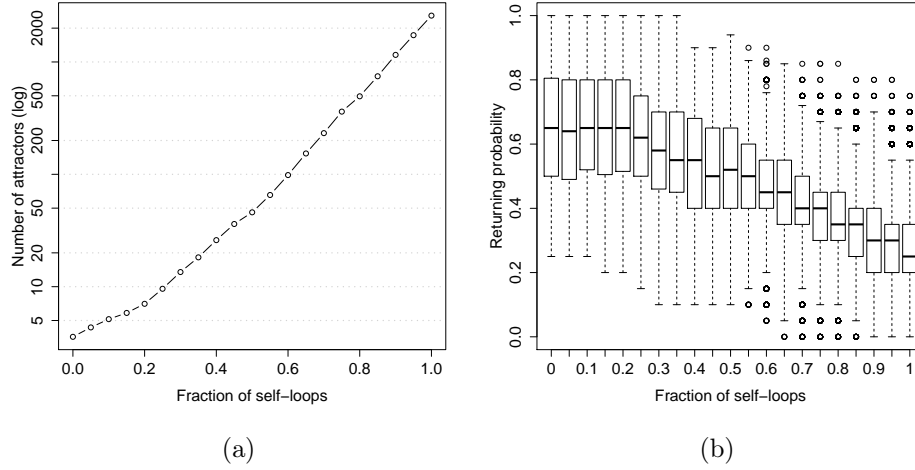


Figure 8.3: Average number of attractors (*i.e.*, both cycles and fixed points) as a function of the fraction of self-loops added with an OR function in RBNs originally with $k = 2$ (8.3a). Distribution of the probabilities of returning to an attractor after one node flip (8.3b).

CONST-RND

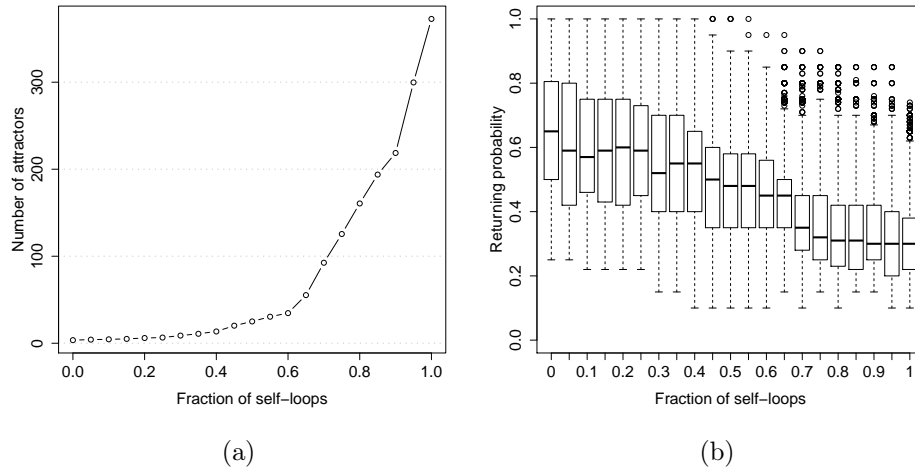


Figure 8.4: Average number of attractors (*i.e.*, both cycles and fixed points) as a function of the fraction of self-loops added in RBNs originally with $k = 2$ (self-loops are introduced by rewiring a randomly chosen input) (8.4a). Distribution of the probabilities of returning to an attractor after one node flip (8.4b).

CONST-OR

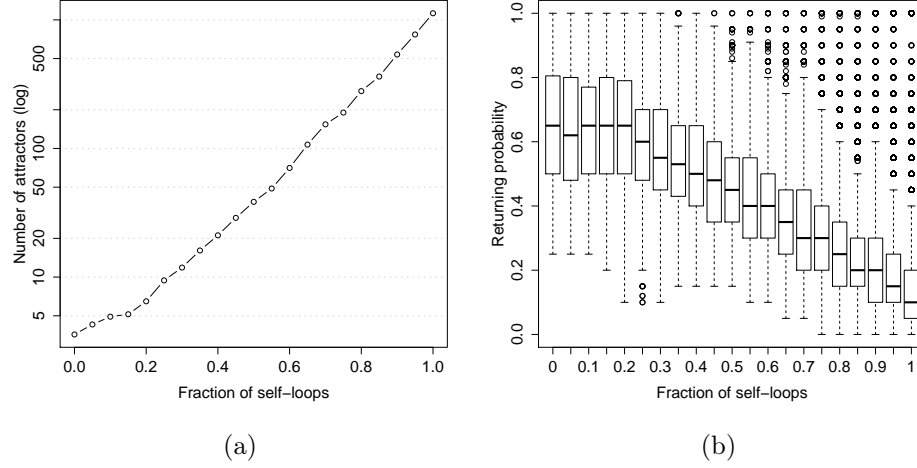


Figure 8.5: Average number of attractors (*i.e.*, both cycles and fixed points) as a function of the fraction of self-loops added in RBNs with $k = 2$ (self-loops are introduced by rewiring a randomly chosen input and substituting the Boolean function with an OR) (8.5a). Distribution of the probabilities of returning to an attractor after one node flip (8.5b).

that in a BN with random topology and all OR functions it is very likely to have two attractors corresponding to two fixed points $S_0 = (00 \dots 0)$ and $S_1 = (11 \dots 1)$, characterised by a basin of attraction of 1 and $2^n - 1$, respectively. Therefore, in the limit case the number of attractors decreases—instead of increasing as in the case with self-loops—and so we expect experimentally. Results of—statistics over 50—experiments are summarised in Figure 8.6.

We observed that the average number of attractors tends to decrease until almost 80% of OR functions within the network. This result is consistent with literature findings: from theoretical results is known that canalising functions move the RBN with $k = 2$ from a critical dynamic regime towards an ordered regime where we observed that the mean number of attractors is one. Thereafter, we observe a final increasing trend. We conjecture that it is due to the growing prevalence of network with two fixed points. In particular with all OR functions we measure an average number of attractors slightly bigger than two. We think that this result is owing to network topology that prevents the signal to be propagated to the whole network.

8.1.5 Discussion and Conclusion

If we want to stay close to Kauffman’s interpretation of attractors, during the process of differentiation a RBN evolves and passes through different at-

ONLY-OR

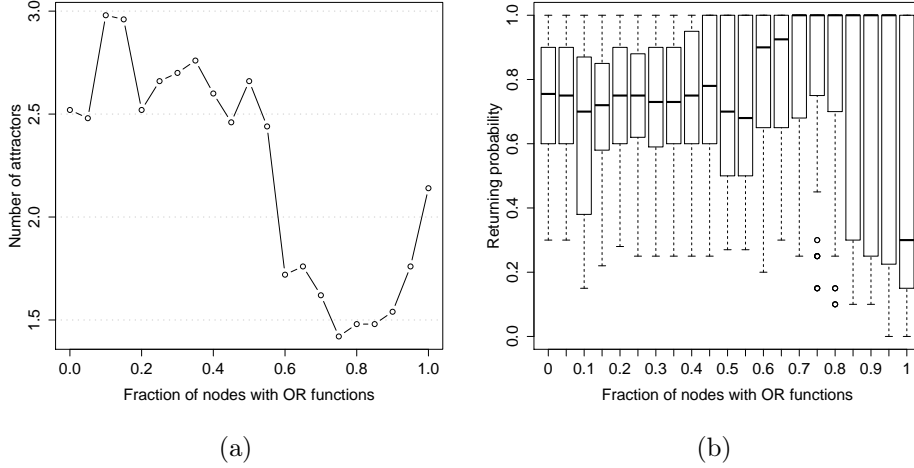


Figure 8.6: Average number of attractors (*i.e.*, both cycles and fixed points) as a function of the number of nodes with an OR function (8.6a). Distribution of the probabilities of returning to an attractor after one node flip (8.6b).

tractors that represent different cellular states, from stem cells to terminally differentiated cells. In this vision, the number of attractors models cellular diversity, while attractor stability models how strong must a signal be to move from one cell type to another. A tight balance between diversity and robustness ensures the perfect homeostasis known in multicellular organisms.

By analysing the impact of positive self-loops in RBN attractor landscape, we observed that they have an important role in network dynamics, and particularly on the number and stability of attractors. On one side they bring diversification, on the other side they seem to be responsible for instability. An operating point, where the balance is perfect as in biological world, is worth to be found.

More than that, biological research identified, for each differentiation state, a set of markers that characterise and identify the differentiation state. In RBNs, within each attractor, only a subset of nodes maintains its state (on/off). We here speculate that these nodes can model the concept of markers—an in-depth analysis of this claim is devoted to future work. In the model presented in this work, self-loops are the mechanism that contributes maintaining the node state. This is particularly true considering those networks where self-loops are introduced with the OR function. There, their role is exactly to keep the local stability on a subset of nodes, representing the marker genes. If the network is in the operating point, “some self-loops but not too many”, they have the crucial role to cause diversification, *i.e.*, differ-

ent cell types, without harming attractors stability, *i.e.*, cell type robustness.

Finding this operating point is thus crucial. However not trivial, especially because homeostasis in multicellular organisms is the result of a number of different mechanisms. This means that, including in the model also other phenomena and bricks, can change the dynamic described in this work. In particular we draw the reader attention to auto-inhibitory processes, epigenetics and cell-to-cell interactions. Auto-inhibition negatively regulate gene expression. Epigenetics affects gene expression by changing the chromatin accessibility. As a consequence, cells with the same set of genes respond differently to the same signal. Cellular interactions influence the intracellular GRN dynamic by means of signals crosscutting cell membranes and whose effect is typically to activate or inhibit the expression of the target gene.

We conclude that results shown in this work suggest to study the advantage of having self-loops in genetic networks during differentiation processes. However, further investigation is necessary to provide a more complete analysis and understanding of the results observed in this work, supporting our findings with theoretical verification and estimation. Moreover, future work will be devoted to investigate what mechanisms counterbalance the effect of self-loops on attractor robustness, which is, as discussed previously, a fundamental property for modelling cell dynamics.

Finally, as mentioned in the Introduction section, we can think that this catalogue of bricks we are building, and whose functions we are analysing, are elementary building blocks that can be combined to face the reverse engineering problem of reconstructing real GRNs or designing GRN model with desired dynamics properties for artificial purposes. In addition, this approach can give us insights of the evolutionary processes that biological GRNs have undergone.

8.2 Self-loops favour diversification and asymmetric transitions between attractors

Changes in network topology and functions strongly impact characteristics of Boolean networks attractor landscape. With this contribution, we studied how self-loops influence diversified robustness and asymmetry of transitions. The purpose of this study is to identify the best configuration for a network owning these properties.

Our results show that a moderate amount of self-loops make random Boolean networks more suitable to reproduce differentiation phenomena. This is a further evidence that self-loops play an important role in genetic regulatory networks.

8.2.1 Motivation and goal

As we have repeatedly stressed in previous chapters, Boolean network (BN) models provide a suitable generic model for cell differentiation [Villani et al., 2011, Furusawa and Kaneko, 2012, Huang et al., 2009b]. The main assumption in this modelling perspective is that attractors, or sets of attractors, represent cell types. In multicellular systems differentiation is characterised by differential expression of genes, meaning that each cell type expresses only a subset of genes called *markers*. In the same way, at each attractor of the BN corresponds the dynamic activation of only a subset of nodes. Accordingly, transitions between attractors epitomise cell differentiation stages that bring changes in the pattern of active/inactive genes. During the process of cell differentiation, cell responds differently to external cues, such as epigenetic modifications. By that we mean robustness of cell states being not the same during the whole process of differentiation. Zhou et al. [Zhou et al., 2016] calls this property of gene regulatory network as *relative stability*. In this view, for a network to be a suitable model for cell differentiation, one would require to have attractors characterised by different degrees of robustness, such that some of them are more responsive to external stimuli and perturbations, while others are rather insensitive to external perturbations and so more stable. This property can be expressed as *diversification in attractor robustness*. In addition, as pointed out by Zhou et al. [Zhou et al., 2016] in a recent work on Boolean models for pancreas cell differentiation, a further requirement of the model is to be characterised by *asymmetric transition probabilities between attractors*. This further property accounts for a preferential directionality of the differentiation process, that anyway does not exclude reversibility. These two dynamical properties are the combined result of topological and functional settings of the network model instance. It is therefore important to identify specific settings that favour the arising of such properties. With this work we tackle this issue addressing the question as to whether self-loops in Boolean network models may positively contribute to attaining dynamics with diversified attractor robustness and asymmetric transitions.

Results show that, *ceteris paribus*, networks in which few nodes have self-loop are more likely to exhibit the properties mentioned above.

8.2.2 Methods

Attractors in BNs are unstable with respect to perturbations (i.e. temporary node value flips), therefore after a node flip the trajectory either returns to the same attractor or it reaches another one. Attractor transition probabilities can be computed on the basis of the perturbation mechanism adopted. In this work we suppose that only one node at a time can be perturbed and that only states belonging to an attractor can be subject of such perturba-

tion. In practice, we apply a logic negation to each node of each state of each attractor in turn and we check in which attractor the dynamics relaxes. This hypothesis is based on the assumption that perturbations are non frequent and so the probability of affecting more than one node at a time is negligible, and the same holds for perturbations occurring during transients, which usually occupy a tiny fraction of time with respect to attractors along BN trajectory. As shown in [Braccini et al., 2018], these hypotheses are quite loose and results obtained in this setting are comparable to stochastic simulation of perturbed BNs. Under these assumptions, the probability of a transition between attractor A and attractor B can be computed in principle by taking the frequency of transitions between A and B among all the possible node flips along attractor A . When networks are large, we often resort to sampling instead of enumerating all the possibilities. The probability transition matrix is usually named *Attractor Transition Matrix* (ATM) 6.5.1. The diagonal of the ATM account for the robustness of attractors, as diagonal values represent the probability of returning to the same attractor after one flip. Moreover, in general, transition probabilities are not symmetric, i.e. $p(A \rightarrow B) \neq p(B \rightarrow A)$. We observe that its diagonal values may be correlated with the attractor basins, but it is important to note that a high value in the ATM diagonal does not necessarily correspond to an attractor with a large basin of attraction, because the values in the ATM are computed by considering single perturbations occurring along the attractor states, while the attractor basin is defined in terms of a fraction of the entire state space.

8.2.3 Diversification and Asymmetry

A BN suitable to represent a differentiation process should exhibit different degrees of robustness, in the same way as cells at different differentiation stages are more or less sensitive to external perturbations. This property can be evaluated by quantifying the different values along the ATM diagonal and their range. Furthermore, in a recent work [Zhou et al., 2016], Zhou and collaborators add another important requirement: the dynamics among attractors should be asymmetric, i.e. $p(A \rightarrow B)$ and $p(B \rightarrow A)$ should be different enough to observe a significant degree of irreversibility. This property can be evaluated by quantifying the distribution of values along the ATM rows.

To quantitatively evaluate these two properties, we defined two functions. They are not meant to be a formal definition for the properties themselves, but in our view they provide a good quantitative approximation of their quality and make it possible to compare different network configurations. Given the ATM, whose values are denoted by $T_{i,j}$ —where i is the index of rows, and j the index for columns—we order rows so that values on the main diagonal are in ascending order. The matrix is by definition square of size m , where m is the number of attractors. In the following, with ATM, we

will refer to this sorted matrix.

Diversification in attractor robustness is estimated on the basis of the following function:

$$f_1 = \sum_{i=2}^m (T_{i,i} - T_{i-1,i-1}) = T_{m,m} - T_{1,1} \quad (8.1)$$

This function estimates the range of attractor robustness. We observe that $0 \leq f_1 \leq 1$, but in general—and mainly in random models—it decreases with m as the more are the options to escape from an attractor, the lower the probability of returning to it after a perturbation.

Asymmetry in transition probabilities is defined in terms of the sum of transition probabilities in the triangle above (Q_a) and below (Q_b) the main diagonal:

$$f_2 = Q_a - Q_b = \sum_{i=1, j=i+1}^{m-1} T_{i,j} - \sum_{i=2, j=i-1}^m T_{i,j} \quad (8.2)$$

The intuition behind this definition is that high values of f_2 characterise asymmetric transitions between attractors. Moreover, as the transition matrix is sorted by main diagonal values, the function estimates the extent to which transitions from less robust to more robust attractors are favoured. In principle, f_2 may range in $[-m+1, m-1]$ but the actual distribution of values strongly depends upon the models used.

In general, we may assume that the higher f_1 and f_2 computed on an ATM of a Boolean network the higher the potential of that network to model a differentiation process.

8.2.4 Self-loops

Network topology and functions impact the BN attractors landscape [Ahnert and Fink, 2016]. In the Section 8.1 we have reported the obtained results related to the impact of self-loops in RBNs [Montagna et al., 2018]. A self-loop in graph theory is defined as an arc that connects a vertex to itself, and in the context of BNs this implies that the state of a node with a self-loop at time t depends also on the state of the very same node at time $t-1$. There, we have shown that a major effect of incrementally adding self-loops in RBNs is that (i) the number of attractors increases (in some cases exponentially) and (ii) the probability of returning to the same attractor after a perturbation decreases with the fraction of nodes with self-loops. In fact, this last effect might be detrimental to cell dynamic if generalised to all attractors. Therefore, as self-loops have been observed in genetic networks reconstructed from real data, this effect should be limited to few attractors and anyway be compensated by other positive effects, maybe concerning cell differentiation itself. In this work we investigate the impact of self-loops on

the properties mentioned above, with a twofold aim: identify possible structural network characteristics that favour their use as differentiation models and find support to the appearance of self-dependencies in biological genes.

8.2.5 Results

We initially sampled RBNs with $n = 15$, $k = 2$ and $p = 0.5$ —following the long standing hypothesis according to which real GRNs operate in a critical regime (or ordered but not chaotic) [Shmulevich et al., 2005]—subsequently modified with self-loops according to the four different schemas already described in Section 8.1.2.

In all the listed cases, the choice of the incoming link to be substituted and the node to which to add the self-loop was performed in a random fashion, choosing with uniform probability among the nodes. For computational reasons, we limited both the maximal number of nodes with self-loops and of nodes in the BNs so as to make feasible the exhaustive exploration of BNs' state space, as well as the perturbations for generating the ATM. Statistics on larger BNs, requiring a sampling of the possible initial states and perturbations, are subject of future work.

Note that in our models there is actually no semantics associated to 0 and 1. However, since for the sake of simplicity we chose a specific canalising function, the OR function, this implies that if 1 is associated with the active state of the node, OR acts as a canalising activating function—and, clearly, OR with self-loops means self-activation. As originally introduced by Kauffman [Kauffman, 1993], a canalising function is a Boolean function in which there exists an input value that fully determines the output value, regardless the values of other inputs. To be coherent in the discussion of the results, we stick to this choice and hereafter we designate 1 as the active state. Anyway, the fact that we observe a specific effect of OR functions just means that this effect can be achieved by any canalising function of this kind. Moreover, as we have shown in [Montagna et al., 2018], the effect is produced by the combination of self-loops and OR functions together, rather than being the consequence of an increased fraction of OR functions alone. As f_1 and f_2 depend on the number of attractors and in an ongoing work we have observed that the distribution of the number of attractors is not uniform among RBNs of a given size, in light of this we have sampled the BNs space until the desired fixed number of nets had reached and then the statistics we computed are based on the same number of nets (30) per number of attractors, from 1 to 20—hence, statistics are computed over 600 networks. Indeed, we believe that the advantage of self-loops should be observed *ceteris paribus*, i.e. evolutionarily speaking, once the network has reached its minimal configuration in terms of attractors. In addition, when BNs are used for modelling real cells, the first step is usually to configure a network with a given number of attractors (cell types), as done for example

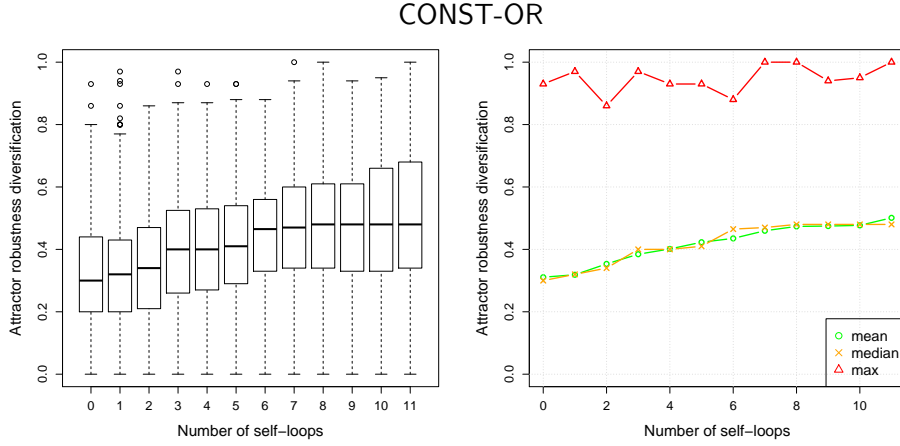


Figure 8.7: Boxplot (left) and main statistics of f_1 (right), in the case with $k = 2$ and self-loops combined in OR.

in [Zhou et al., 2016]. Results are shown by means of boxplots, so as to have a visual reckon of the distribution, and a plot with maximum, mean and median (the minimum value is omitted as it is the same for all the statistics for each function). On the x -axis we find the number of self-loops and on the y -axis either f_1 or f_2 . Generic statistics on attractors are omitted because not relevant for this contribution; however, the interested reader can find such statistics in our previous work [Montagna et al., 2018]. An in-depth analysis of attractors number distribution in this class of BNs is subject of ongoing work.

Results concerning robustness diversification (measured by means of f_1) are summarised in Figures 8.7–8.10. We first note that, in all the cases, mean and median are almost overlapping, meaning that the distribution is symmetrical. Moreover, as the maximum (shown also in the right plots for clarity) is always much higher than the mean and follows approximately the same shape, we can conclude that the distribution is quite wide—a property that favours evolutionary processes. We observe that in those cases in which self-loops play a self-activating function (OR cases), the diversification steadily increases (CONST-OR) or reaches a maximum at a moderate amount of nodes with self-loops (AUGM-OR). We observe that, when the self-loop may play any functional role – i.e. when it is combined in a random function – this effect is limited, if not negative.

Results concerning the asymmetry of transitions between attractors are summarised in Figures 8.11–8.14. Results are analogous to the previous case: when self-loops play a self-activating function (OR cases), the asymmetry increases steadily (CONST-OR) or up to a point after which it starts to decrease (AUGM-OR). Conversely, when the self-loop is associated to a random function, the average impact on asymmetry is negligible.

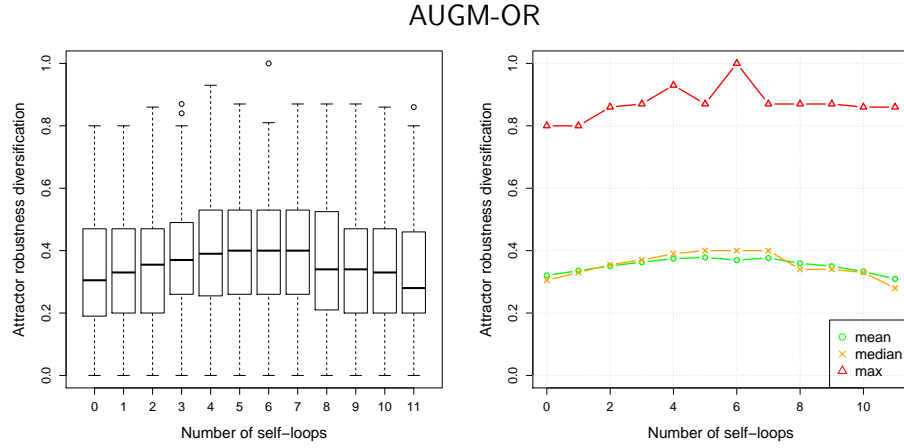


Figure 8.8: Boxplot (left) and main statistics of f_1 (right), in the case with self-loops added to the nodes in OR.

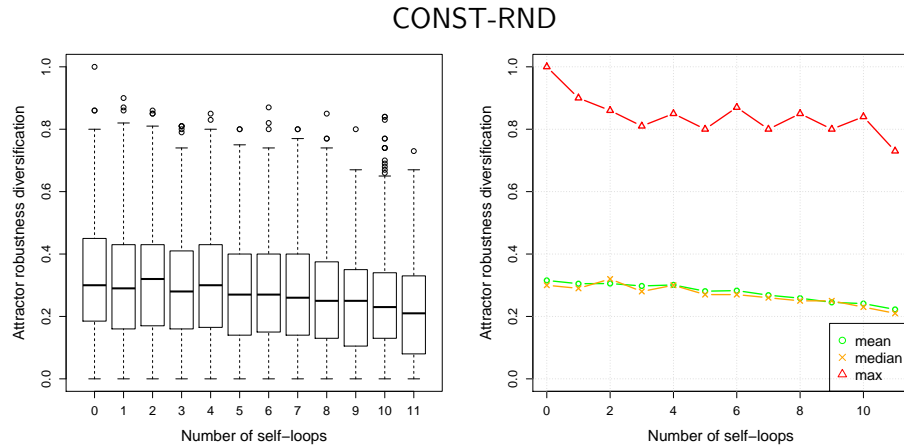


Figure 8.9: Boxplot (left) and main statistics of f_1 (right), in the case with $k = 2$ and self-loops combined as a random function.

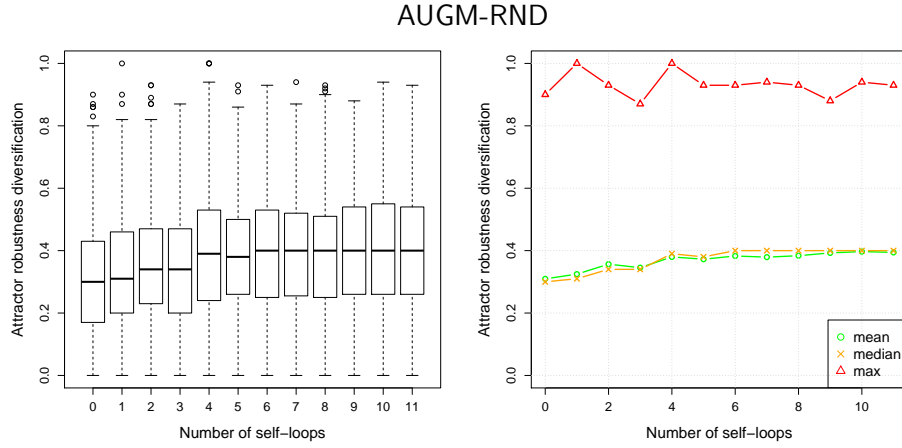


Figure 8.10: Boxplot (left) and main statistics of f_1 (right), in the case with self-loops added to the nodes in a random function.

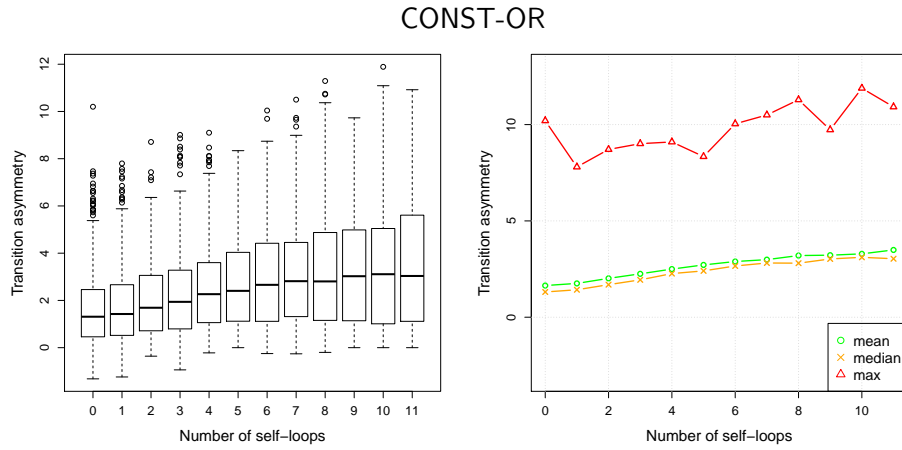


Figure 8.11: Boxplot (left) and main statistics of f_2 (right), in the case with $k = 2$ and self-loops combined in OR.

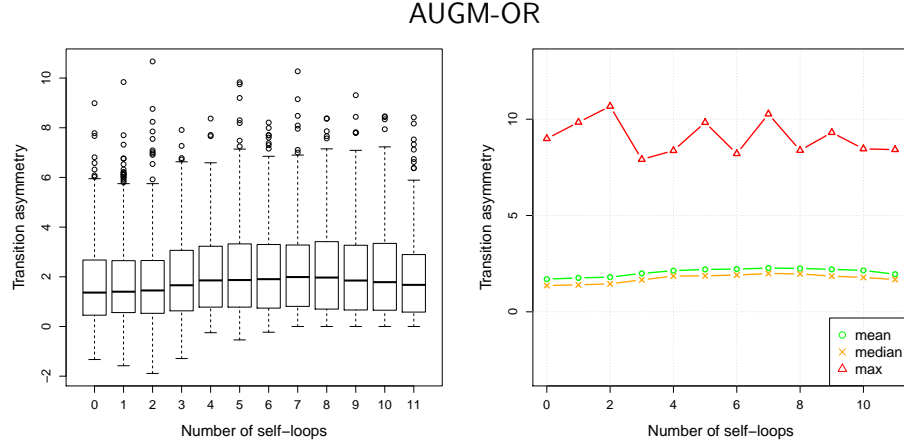


Figure 8.12: Boxplot (left) and main statistics of f_2 (right), in the case with self-loops added to the nodes in OR.

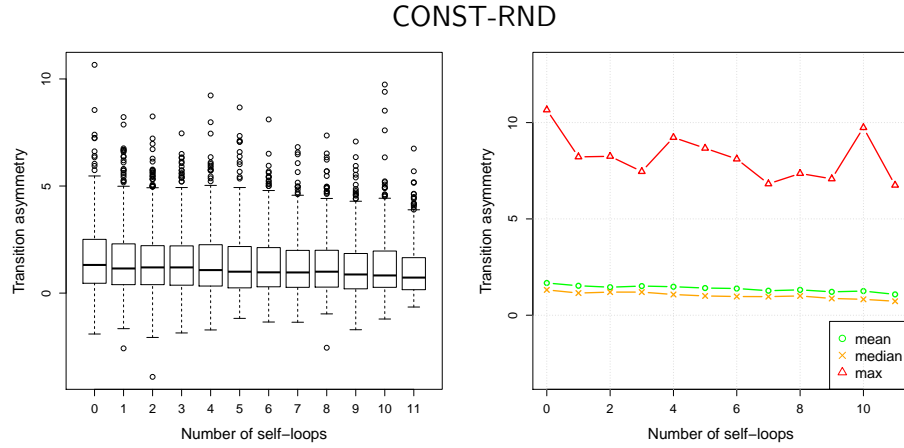


Figure 8.13: Boxplot (left) and main statistics of f_2 (right), in the case with $k = 2$ and self-loops combined as a random function.

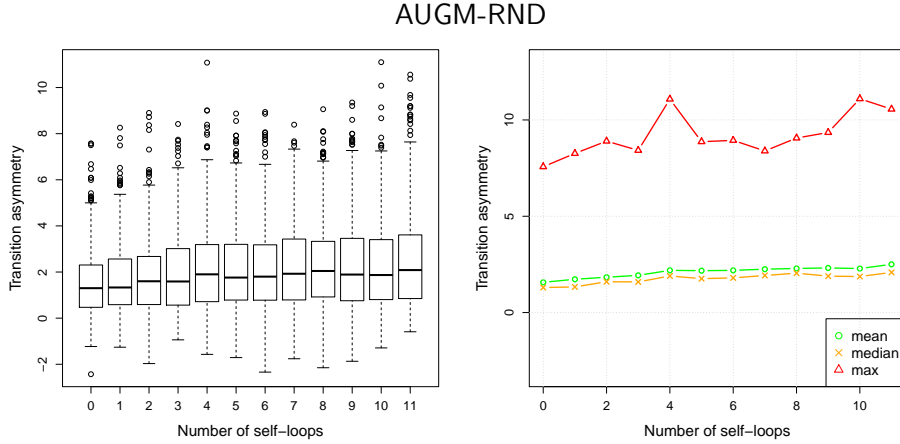


Figure 8.14: Boxplot (left) and main statistics of f_2 (right), in the case with self-loops added to the nodes in a random function.

Summarising, these results support the conjecture that, in BN models, a small fraction of node with self-loops associated to a canalising function may favour both robustness diversification and asymmetry in transition probabilities. This result observed in a computational model is indeed complemented by experimental evidence in real cells where self-activation/inhibition schemes are found in a limited, but not negligible, number of genes [Raj et al., 2010].

8.3 Conclusion

According to the ensemble approach [Kauffman, 2004], it is important to identify structural and functional properties of genetic regulatory network models that make them more suitable for reproducing typical cell dynamics. In this perspective, we have investigated the effect of self-loops on RBNs addressing the question as to whether they may provide a positive contribution to modelling cell differentiation. Attractors in BNs epitomise cell types and transitions between attractors occur when attractor states are perturbed, either by specific signals or perturbations. We have defined two functions as proxies for measuring two prominent properties acknowledged to play an important role in BN models for differentiation: diversification in attractor robustness and attractor transition asymmetry. Results show that a small fraction of canalising self-loops (approximately between 0.25 and 0.3) favours higher values of these functions. This result sheds light on the positive role of self-loops in BNs for modelling cell dynamics. Moreover, by the outcome of this study we contribute answering the question raised in [Montagna et al., 2018] concerning the existence of a specific fraction of self-loops enabling an optimal balance between robustness and flexibility. In addition, this inves-

tigation may provide insights on the evolutionary processes that biological cells have undergone. This contribution is just another step in the direction of detecting structural and functional characteristics (*bricks*) in GRN models that make them suitable for modelling cell differentiation processes. Further goals in our research agenda involve the identification of other bricks to be used as elementary building blocks in the problem of designing GRNs for biological as well as artificial purposes. Moreover, since our measures are based on a particular procedure to assess the attractors' robustness recapitulated by the ATM, is of paramount importance the comparison of this last with other ways to characterise their relative stability, as done in [Joo et al., 2018]. In addition, given the preliminary nature of the work, a comparative study between different metrics (e.g. the Gini index [Gini, 1912]) which could characterise the diversification and asymmetry properties of the attractor landscape have not yet been done. Still, it is already scheduled for future work.

8.4 The impact of self-loops on BN attractor landscape and implications for cell differentiation modelling

In the work [Montagna et al., 2018] presented in Section 8.1, we have found that the impact of self-loops in random Boolean networks is to increase the average number of attractors and reduce their average robustness. This last result is not in agreement with biological networks, which are undoubtedly robust and yet contain self-loops.

With this work we have shed light on this conundrum by investigating in more depth the impact of self-loops in Boolean networks on both the number of attractors and their robustness. We first showed that the number of nodes with a self-loop is indeed positively correlated with the average number of attractors and we provide a formal model for this relation. Subsequently, we showed that, if we restrict statistics to networks with the same number of attractors, the maximal robustness of the attractors increases with the fraction of nodes with a self-loop. In other words, the advantage of self-loops is still observable in Boolean models but by comparing attractor robustness *ceteris paribus*. In addition, our results show that the variability of attractor robustness tends to increase with the fraction of nodes with self-loops. This outcome suggests that direct autoregulation may provide an advantage in the evolution of the basic dynamic mechanisms of cells.

These results provide further support to the use of Boolean networks for modelling cell dynamics (e.g. differentiation processes) and suggest that self-loops have to be taken into account in the *ensemble approach* [Kauffman, 2004], which aims at identifying generic properties so as to match some statistical features of the target biological systems.

8.4.1 Motivation and Goal

Being the asymptotic states of the system, attractors play a prominent role in BNs [Kauffman, 1993, Kauffman, 2000]. In particular, by following the recently proposed dynamical systems view of cell differentiation (see the works [Huang et al., 2005, Furusawa and Kaneko, 2012, Huang et al., 2009b, Villani et al., 2018] and Part II for more details), attractors—or subsets of attractors—represent cell types. Accordingly, transitions between attractors epitomise cell differentiation stages that bring changes in the pattern of active/inactive genes. In this view, for a network to be a viable model for cell differentiation, one would require to have a suitable number of attractors characterised by varying degrees of robustness, so as to be able to reproduce the transitions between cell types.

In [Montagna et al., 2018] we presented a preliminary study that analyses whether adding self-loops in RBNs affects attractor number and type and, possibly, their robustness. We observed that the number of attractors is higher in networks with self-loops and grows quasi-exponentially with the number of self-loops. At the same time, attractor robustness tends to be smaller than in RBNs without self-loops.

These results are not completely coherent with the role self-loops have in biological systems. Indeed, autoregulatory circuits—biological components that are (in)directly influenced by their very product—are pervasive in biological organisms and they are actively involved in conferring mutational, environmental, recombinational, or behavioural robustness. The effects of these circuits manifest themselves as emergent properties on multiple scales, in time (development/evolution), in space (populations) and on different levels of the biological organisation (from molecular up to entire organisms). Buffering of noise and incomplete penetrance [Chalancon et al., 2012], autocatalysis, homeostasis and buffering gene dosage [Thomas et al., 1995], genetic switches—like *Sxl* gene in sex determination of *Drosophila* [Thomas, 2002] and *Cl protein* in lytic or lysogenic phase control in bacteriophage lambda [Crews and Pearson, 2009]—and chromatin mediated autoregulation [Fisher, 2002] are just some of the most prominent examples of the observable effects of positive or negative, direct or indirect autoregulatory circuits. Autoregulation patterns assume particular relevance in transcriptional regulation. Indeed, in the works by Alon and colleagues [Alon, 2006, Alon, 2007, Milo et al., 2002], autoregulatory circuits have been identified in transcription networks as network motifs, i.e. recurring building-block patterns found in complex networks. Just to mention an example of their amount in a real organism, *E. coli* presents 40 transcription factors that regulate the transcription of their own genes, out of a total of 420 transcription factor encoding genes.¹

Particularly noteworthy to the purpose of this work are the functions that

¹according to [Alon, 2006]

positive autoregulations carry out in differentiation, and therefore in development. According to [Alon, 2007], positive (negative) autoregulation occurs when a transcription factor enhances (represses) its own rate of production. It is the memory capacity typical of positive circuits made by maintaining gene expression, and so acting as genetic switches, that makes them important in biological development. Therefore, autoregulation of important developmental regulatory proteins can *lock-in* their expression and so induce the maintaining of attained cell fates or developmental states [Crews and Pearson, 2009]. Thomas in his works [Thomas, 2002, Thomas et al., 1995, Thomas et al., 1976] remarks the key role of autoregulation in the context of cell differentiation firstly by demonstrating that a positive loop is necessary for multistationarity and subsequently, following the hypothesis of Delbrück, that differentiation represents the biological aspect of the latter. In addition, Alon [Alon, 2006] addresses a mathematical systematic study of the mechanisms characterising autoregulation generic dynamical properties: negative autoregulation speeds-up transcription response time, whilst positive ones slow down the transcription factors response time and are able to create bi-stability. Huang et al. [Huang et al., 2007] study the interactions between two key transcription factors in blood differentiation, namely GATA1 and PU.1, to understand the discrete cell fate decisions that multipotent cells undergo during development. These two transcription factors promote the erythroid or myelomonocytic lineage respectively. In a dynamical systems vision of cell differentiation, the authors formulate a minimal mathematical model of the functional interactions of the two above-mentioned transcription factors that in a qualitatively way reproduce—in the GATA1 and PU.1 plane—the observed experimental genome-wide trajectories of the transcriptome during differentiation. In addition to mutual inhibition, auto-stimulation of GATA1 and PU.1 turned out to be fundamental to give rise to the metastable state characterised by the promiscuous expression of both transcription factors and representing the progenitor cells.

In the light of the previous—non-exhaustive—list of biological and modelling examples in which autoregulations play relevant roles in the organism functions, especially in transcription networks, we cannot ignore their roles also in Boolean models of GRN. In fact, Boolean reconstructed models of GRNs, obtained by making use of consolidated biological knowledge of transcription factors interactions or following *ad-hoc* procedures for synthesising models able to reproduce observed data, are characterised by autoregulations. As an example we cite the reconstructed Boolean network representing the core endogenous network of early myeloid cell-fate determination [Su et al., 2017] that presents nearly 10% of nodes with self-loops. In [Joo et al., 2018] the authors have reconstructed a BN model of the control mechanisms that drive the epithelial-to-mesenchymal transition (EMT). In that minimal GRN model autoregulations are necessary to create the required stable attractor states and, in particular, the hybrid cell state that presents

in the gene expression profile both the epithelial and mesenchymal features. Moreover, by analysing the database (<https://cellcollective.org/#>) of Boolean-rule based computational models of large-scale biological networks available at the Cell Collective website [Helikar et al., 2012], we have ascertained that the 54% of networks has self-loops. Nevertheless, a systematic study aimed at identifying **generic properties** and qualitatively characterising the impact of nodes with self-loops in BNs is missing. Indeed, modelling and analysing specific genetic circuits is a valuable and necessary step to pursue an ever-increasing understanding of the mechanisms that underlie biological organisms, but it suffers from limitations because these circuits are not actually functionally independent [Huang et al., 2007]. Therefore, in order to frame up their effects in the dynamics of the complex networks of which they belong, it is necessary to address also this study in terms of generic properties. Our work is in the track of the long-term research aiming at constructing a **synthetic network biology** theory since self-loop may represent a possible **functional bricks**.

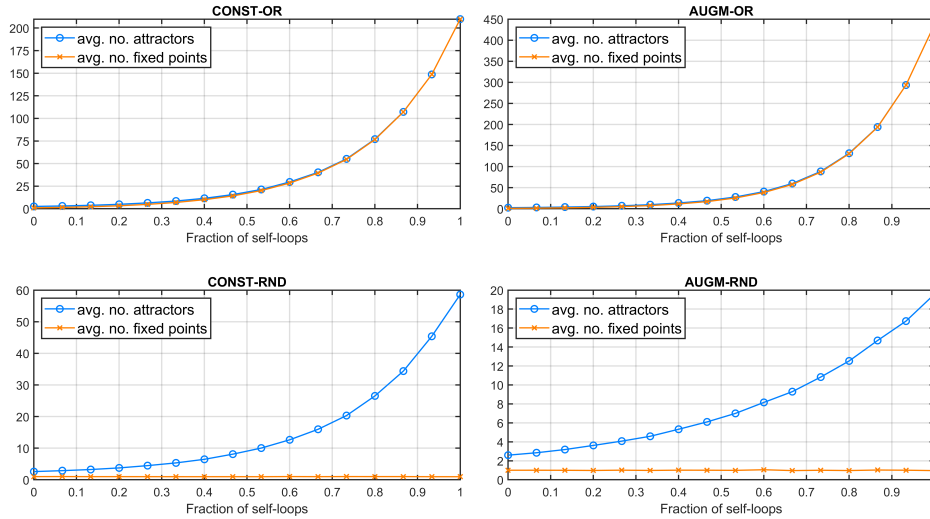


Figure 8.15: Average number of attractors in 2×10^4 RBNs with 15 nodes as a fraction of self-loop varying from 0 to 1. Results are shown for each of the four configurations.

8.4.2 Methods

The experiments we performed concern RBNs in which the main factor we control is the fraction of nodes with a self-loop. Starting from a RBN with n nodes denoted by integer values $V = \{1, 2, \dots, n\}$, k inputs per node and bias b , a self-loop can be introduced in a node either by rewiring an input or by adding a new one. The function of nodes with self-loops may be

arbitrarily altered or left random. Two are the main reasons for investigating the impact of autoregulation as self-loops in random Boolean models: the first is that BNs are among the most used GRN models and a wealth of results on RBNs is already available, therefore we can compare our findings against an established and well known literature. The second reason is that if self-loops make RBNs somehow more adapt to model differentiation processes, then they should be taken into account in artificial evolution experiments and RBNs with self-loops may provide a promising initial condition for such studies.

We performed the experiments according to the following experimental setting. Initial RBNs were created with $k = 2$ and function bias b equal to 0.5—the value of these parameters grounds on biological plausibility [Kauffman, 1969b, Kauffman, 1969a, Serra et al., 2004b]. These networks have been modified by selecting at random n_s nodes in which rewiring or adding a link in self-loop. For brevity, let's denote by $V_s \subseteq V$ the nodes with a self-loop. When self-loops are introduced by rewiring, all nodes in V have exactly k inputs; conversely, when self-loops are added, the distribution of node in-degree changes as n_s nodes out of n have $k + 1$ inputs. We decided to test both variants so to have a wider picture of the effects of self-loops inserted in RBNs. Besides adding self-loops, a decision has to be taken concerning the functions of nodes in V_s : since we started from RBNs, the function may still be random (with bias kept to b also in the case $k + 1$) or it can be set to a specific Boolean function. This latter case is the most relevant for GRN modelling, as self-loops found in biological cells have usually a canalising role [Raj et al., 2010] instead of playing any function. As originally introduced by Kauffman [Kauffman, 1993], a canalising function is a Boolean function in which there exists an input value that fully determines the output value, regardless of the values of other inputs. We then chose to test also the case in which the Boolean function of nodes in V_s is either logical OR between the value of the node itself and the other input (k constant case) or the OR between node value and the random function initially set, for the $k + 1$ case. Formally, let i be a node in V_s and f_i the original function; then, the new Boolean function \hat{f}_i is defined as follows:

- case k constant: $\hat{f}_i = x_i \vee x_j$, where j is the other input of i ;
- case $k + 1$: $\hat{f}_i = x_i \vee f_i$.

Note that in BN models there is actually no semantics associated to 0 and 1. However, since for the sake of simplicity we chose one specific canalising function—the logical OR—this implies that if 1 is associated with the active state of the node, OR acts as a canalising activating function—and, clearly, OR with self-loops means self-activation. Anyway, if we observe a specific effect of the OR function, it just means that this effect can be achieved by any canalising function of this kind. Our main interest is indeed on this

canalising functional role, whilst the case with random functions is kept just for comparison. The set of BN variants used in the experiments we performed are described in Section 8.1.2.

Attractors in BNs are unstable with respect to perturbations (i.e. temporary node value flips), therefore after a node flip the trajectory either returns to the same attractor or it reaches another one [Ribeiro and Kauffman, 2007, Villani and Serra, 2013b]. Attractor transition probabilities are computed by exerting perturbations to each attractor and the probability of returning to an attractor after a perturbation is taken as an estimation of *attractor robustness*. According to [Jen, 2003], robustness is a broader concept than stability—which is a well-defined mathematical notion in dynamical systems—as it is related to *feature persistence* under a wider spectrum of perturbations of different nature. The notion of attractor robustness used in this work, which is related to the concept of *robust adaptation* defined by Kitano in [Kitano, 2004], is not limited to determine single attractor’s stability since it provides a quantitative measure, *i.e.*, the probability of returning to the same attractor.

This metric is clearly a function of the kind of perturbation exerted. In this work we suppose that only one node at a time can be perturbed and that only states belonging to an attractor can be subject to such a perturbation. This approach is common in dynamical systems, in which stability is indeed evaluated in stationary states. Moreover, the single flip hypothesis is based on the assumption that perturbations are not frequent with respect to network updates, so the probability of affecting more than one node at a time is negligible; the same consideration holds for perturbations occurring during transients, which typically last a tiny fraction of time with respect to attractors along BN trajectories. This assumption is reasonable in particular when BNs are used to model cell dynamics [Serra et al., 2010, Villani et al., 2011, Villani and Serra, 2013b].

In practice, we apply a logic negation to each node of each state of each attractor in turn and we check in which attractor the dynamics relaxes. The probability of a transition between attractor A and attractor B is computed by taking the frequency of transitions between A and B among all the possible node flips along attractor A .² As shown in [Braccini et al., 2018], the results obtained in this setting are equivalent to stochastic simulation of perturbed BNs. This procedure produces the probability transition matrix called ATM 6.5.1. As already stated, the diagonal of the ATM accounts for attractor robustness, as diagonal values represent the probability of returning to the same attractor after a perturbation. To get a more accurate evaluation of attractor robustness, in this work we focus on minimum and maximum values of the ATM diagonal, rather than one single statistics such

²For large size networks one has to resort to sampling, instead of enumerating all the possibilities.

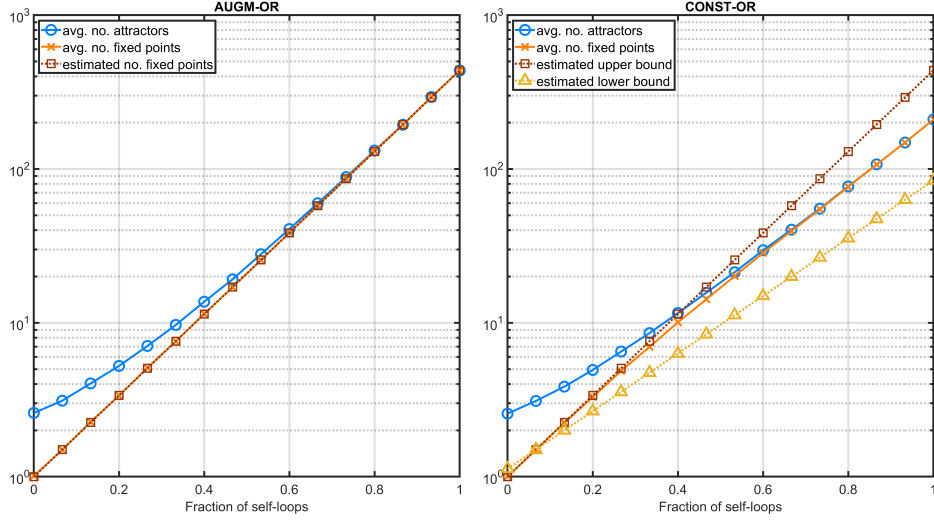


Figure 8.16: Comparison between the mathematical model and experimental results. Self-loop in OR inserted by adding an input (left), and by rewiring one input (right). Logarithmic scale for the y axis.

as the average that might obfuscate the actual features of the distribution. To this purpose we define the two following variables:

$$\delta_{min} = \min \text{diag}(ATM) \quad (8.3)$$

$$\delta_{max} = \max \text{diag}(ATM) \quad (8.4)$$

8.4.3 Results

In this section, we will first show the results on the number of attractors. Subsequently, we will analyse statistics on attractor robustness.

We run experiments for each of the four configurations described in Section 8.1.2. In this work, we aim at providing a detailed picture of the impact of self-loops on BNs and so we choose completeness over statistics in the large. Therefore, the number of nodes n is set to 15 so as to be able to perform an exact computation of the ATM for every possible number of nodes with self-loops $n_s \in \{0, 1, \dots, n\}$. The outcome of our study can be anyway generalised to large size networks. However, small and medium-size networks are often used to model the relations among a limited number of genes, related to a specific function (e.g., the hematopoietic cell differentiation). RBNs are generated with $k = 2$ and bias $b = 0.5$. For each value of n_s and for each variant, we took statistics across $2 * 10^4$ randomly sampled RBNs.

8.4.4 Average number of attractors as a function of self-loops

In this section we show how the average number of attractors is affected by self-loops. Figure 8.15 is composed of four plots, each referring to one of the four configurations in Section 8.1.2. In each graph, the average number of attractors is plotted as a function of the fraction of nodes with self-loops, along with the average number of fixed points. The outcome of these experiments confirms the results presented in our previous work [Montagna et al., 2018]: the average number of attractors increases with n_s/n . Curves follow the same trend in the four plots and are well approximated by an exponential function. However, we can observe two main differences between OR (canalising functions) and RND functions:

1. The number of attractors increases much more in networks with self-loops in OR than in those with RND functions. The difference is roughly of an order of magnitude: we can observe that with OR they vary from an average value of 2.59 in all the networks without self-loops until averages of 209.8 (CONST-OR), 436.9 (AUGM-OR), 58.6 (CONST-RND) and 19.6 (AUGM-RND) in networks with 15 self-loops.
2. For CONST-OR and AUGM-OR, almost all attractors are fixed points even at low values of n_s/n ; conversely, in the RND cases, is the average number of cyclic attractors that grows with n_s/n .

An analytical model

To generalising the previous results and being able to make predictions for any value of n , n_s and bias b , we complemented this analysis with a theoretical estimation of the average number of attractors as a function of these parameters. The model we provide in this section is related to the OR cases, as they are more significant for biological cell modelling than the RND ones. As previously observed, in the OR cases even for few nodes with self-loops almost all attractors are fixed points. Therefore, as we want to have a generalisation for any n , we estimate the number of attractors in terms of number of fixed points. For ease of the proof, we first focus on AUGM-OR and we subsequently modify the model for the CONST-OR case. Our goal is then to estimate the probability that a randomly chosen state $s = (x_1, x_2, \dots, x_n)$ is a fixed point in the case in which self-loops are added to nodes. In the following, we use $f_i(\cdot)$ to denote the application of the function of node i to its inputs values.

Hence we want to estimate:

$$p^*(s) = P\{s \text{ is a fixed point} \mid s \text{ is randomly chosen}\} \quad (8.5)$$

s is a fixed point iff $s = F(s)$, i.e. $(x_1, x_2, \dots, x_n) = (f_1(\cdot), f_2(\cdot), \dots, f_n(\cdot))$.

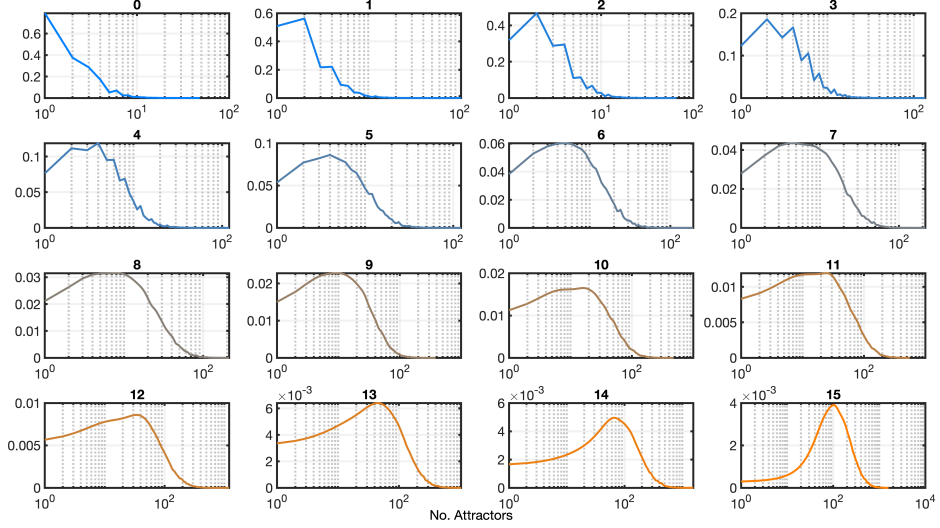


Figure 8.17: Density function of the number of attractors for the CONST-OR configuration. It has been computed from experimental data on 2×10^4 runs with same n_s . The sixteen plots show how it changes with $n_s \in \{0, 1, \dots, 15\}$ (the number of self-loops n_s is reported in bold letters in the title of each plot). Logarithmic scale for the x axis.

Let us focus on node i with two external inputs corresponding to a Boolean function with bias b and a self-loop in OR and estimate the probability $p_i^* = P\{x_i = f_i(\cdot)\}$. We have two cases: (a) $x_i = 1$ and (b) $x_i = 0$.

a) $P\{x_i = 1 \wedge f_i(\cdot) = 1\} = P\{f_i(\cdot) = 1 \mid x_i = 1\} P\{x_i = 1\} = 1 \cdot q = q$, where q is the probability of assigning 1 to value x_i .

b) $P\{x_i = 0 \wedge f_i(\cdot) = 0\} = P\{f_i(\cdot) = 0 \mid x_i = 0\} P\{x_i = 0\} = (1-b)(1-q)$, because b is the probability that—on average— $f(\cdot) = 1$.

Hence, $p^*(x_i) = q + (1-b)(1-q)$. We suppose that initial states are randomly chosen, thus $q = \frac{1}{2}$. Moreover, in our experiments we have $b = \frac{1}{2}$, therefore $p^*(x_i) = \frac{3}{4}$. For a node without self-loops we apply an analogous argument and obtain $\bar{p}^*(x_i) = bq + (1-b)(1-q)$; in our experiments $\bar{p}^*(x_i) = \frac{1}{2}$. Finally, we can derive a formula for the probability of a fixed point of a network with n nodes and n_s nodes with a self-loop added in OR under the hypothesis that all p_i^* are independent:

$$p^*(s) = [q + (1-b)(1-q)]^{n_s} [bq + (1-b)(1-q)]^{(n-n_s)} \quad (8.6)$$

For $b = \frac{1}{2} = q$ we have $p^*(s) = (\frac{3}{4})^{n_s} (\frac{1}{2})^{n-n_s}$.

The comparison between the theoretical value of fixed points and its experimental estimation—based on the statistics on BNs we have performed from

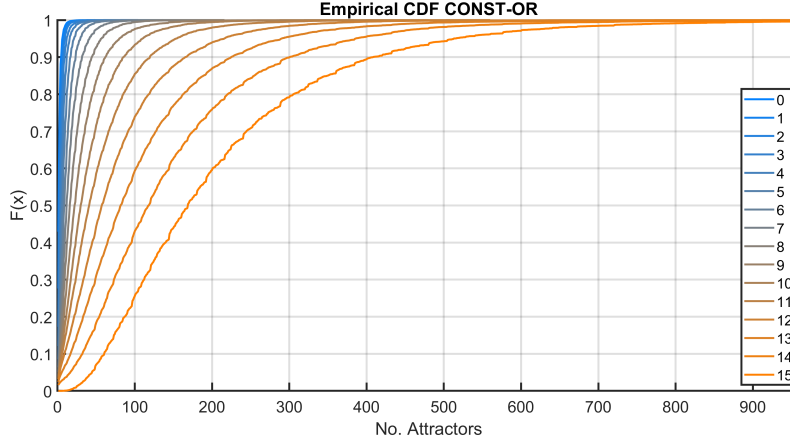


Figure 8.18: Empirical cumulative distribution function (ECDF) of the number of attractors for the CONST-OR configuration. It has been computed from experimental data on $2 \cdot 10^4$ runs with same n_s , with $n_s \in \{0, 1, \dots, 15\}$ (see legend).

simulations of RBNs—is shown in Figure 8.16 (left). The model predicts with high precision the number of fixed points, which is a good estimation of the overall number of attractors. The model slightly overestimates the number of fixed points because we suppose that nodes are independent—in fact, the functional dependence among nodes might rule out some configurations that can, in principle, be fixed points. However, this discrepancy is negligible in this case.

Following a similar reasoning we can derive an analogous formula for the CONST-OR case, in which self-loops are inserted by rewiring an incoming arc and substituting the Boolean function with an OR. In this case we have:

$$p^*(s) = [q + \frac{1}{2}(1-q)]^{n_s} [bq + (1-b)(1-q)]^{(n-n_s)} \quad (8.7)$$

For $b = \frac{1}{2} = q$ we have $p^*(s) = (\frac{3}{4})^{n_s} \frac{1}{2}^{(n-n_s)}$.

The constant value $\frac{1}{2}$ in Equation 8.7 represents the probability that—on average—variable x_i with value 0 does not change its value after the application of the OR function.

The comparison between the theoretical value of fixed points and its experimental estimation is shown in Figure 8.16 (right, dotted line with squares). As we can observe, in this case the hypothesis of independence introduces an error, especially when the number of nodes with self-loop is high. For this case we should apply the chain rule for computing the conjunct probability that every node is constant. Let us denote by $\overline{x_i}$ the event that x_i does not change its value after the update. Thus we have: $p^*(s) = P\{\overline{x_1} \wedge \overline{x_2} \wedge \dots \wedge \overline{x_n}\} =$

$$P\{\overline{x_1}\} P\{\overline{x_2} \mid \overline{x_1}\} P\{\overline{x_3} \mid \overline{x_2} \wedge \overline{x_1}\} \cdots P\{\overline{x_n} \mid \overline{x_{n-1}} \wedge \cdots \wedge \overline{x_1}\}.$$

The dependence among nodes can be simplified because in this topology the nodes with self-loop depend only on one other node as in a ring topology. Hence: $p^*(s) \approx P\{\overline{x_1}\} P\{\overline{x_2} \mid \overline{x_1}\}^{n_s-1} = \frac{3}{4} (\frac{2}{3})^{n_s-1} (\frac{1}{2})^{n-n_s}$, as in the most constrained case x_2 depends on x_1 and, in this case, x_2 is constant in two out of the three cases in which x_1 is constant. In this way, we can provide also a lower bound on the number of fixed points, as depicted in Figure 8.16 (right).

To generalise analytical model results, Figures A.17 and A.18—present in the appendix A—report model predictions varying the bias for the analysed topological OR configurations (CONST-OR, AUGM-OR). We can observe that in the CONST-OR configuration the bias value does not impact the model prediction, while in the AUGM-OR case predictions change with bias values.

As a final note, we observe that the model indeed capture an exponential relation between the average number of attractors and the fraction of nodes with self-loops, as empirically noted in Figure 8.15.

For the sake of completeness, the analytical model for the estimation of the average number of fixed points for the RND cases can be reduced to the following formula $[bq + (1 - b)(1 - q)]^n$ since all nodes have random functions we can not make a distinction between nodes with a self-loop n_s and those without. Since the average outcome of the Boolean functions in RBN ensembles of random functions follows the bias parameter, in Figure A.16 we report the theoretical estimations of the average number of fixed points varying bias values for the RND cases (CONST-RND, AUGM-RND). Noteworthy it is that for both the RND cases the theoretical average number of fixed points is 1—regardless of the bias—and the experimental value reported in Figure 8.15 is perfectly in agreement with it.

8.4.5 Distribution of attractor number

To providing a more detailed picture of the overall attractor number trend we analysed the distribution of the number of attractors of the BNs across the $2 \cdot 10^4$ networks with same configuration and number of self-loops. Moreover, this analysis is useful to study attractor robustness, which is influenced by the number of attractors, i.e. by the ATM size: in BNs a higher number of attractors is likely to correspond to a lower probability of returning to the same attractor after a perturbation, as on average the more the attractors, the smaller their basin of attraction.³ For the sake of brevity, we discuss here the case of CONST-OR, and we refer to the appendix A for additional data and results on the other three configurations.

³In fact, this is a rough reckon that holds on average, because the values in the ATM are computed by considering single perturbations occurring along attractor states, while the attractor basin is defined in terms of a fraction of the entire state space.

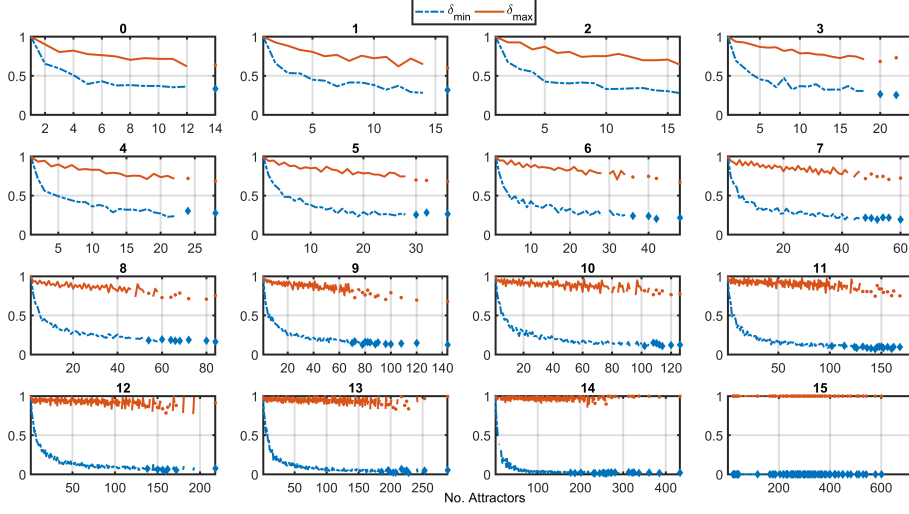


Figure 8.19: δ_{min} and δ_{max} as a function of the number of attractors m for the CONST-OR configuration. They have been computed from experimental data on $2 * 10^4$ runs with same n_s . The sixteen plots show how they change with $n_s \in \{0, 1, \dots, 15\}$ (the number of self-loops n_s is reported in bold letters in the title of each plot). x axis limits change with n_s from 1 to the largest number of attractors for which at least 30 sampled BNs have been found.

Figure 8.17 and Figure 8.18 show the density and the empirical cumulative distribution of the number of attractors, respectively, for each possible value of the number of self-loops in the network.

Density functions in plots of Figure 8.17 are computed with the algorithm implemented in `ksdensity` of Matlab (R2018a). The function returns a probability density estimate for the vector data containing the attractor number from our $2 * 10^4$ experimental results. The result is somehow surprising: for each network setting with at least one self-loop, the density is non-monotonic and a peak can be identified; in other terms, the probability of randomly sampling a network with a given number of attractors is not uniformly distributed, but rather most networks have a number of attractors close to the value corresponding to the maximal density. This peak varies with the number of nodes with self-loops. For instance, in networks without self-loops the value of the density function for a landscape with only one attractor is 0.79, while, once rewiring one node input with a self-loop, two attractors constitute the most common landscape and we have a value of 0.56. Peak value and position for the 16 experiments are summarised in Table 8.1. Peaks of density functions move right as self-loops are added to the network: the probability to have few attractors decreases while the probability to have a large number of attractors increases. Moreover, the peak

Table 8.1: Density function max values and corresponding position, i.e. number of attractors.

n_s	density fun. max value	$x \equiv$ no. attractors
no self-loops	0.7904	1
1 self-loop	0.5619	2
2 self-loops	0.4673	2
3 self-loops	0.1857	2
4 self-loops	0.1187	4
5 self-loops	0.0861	4
6 self-loops	0.0603	5
7 self-loops	0.0435	4
8 self-loops	0.0315	8
9 self-loops	0.0228	10
10 self-loops	0.0165	16
11 self-loops	0.0119	25
12 self-loops	0.0086	34
13 self-loops	0.0064	45
14 self-loops	0.0049	67
15 self-loops	0.0039	102

Table 8.2: Max number of attractors as a function of self-loops n_s

n_s	0	1	2	3	4	5	6	7
A_{max}	48	104	66	136	120	123	192	216
n_s	8	9	10	11	12	13	14	15
A_{max}	212	384	480	560	1200	1050	1545	1601

gets lower while tails of density function get higher and longer. Generally speaking the peak moves to the right while it drops down. Accordingly, we extracted the maximum number of attractors observed (A_{max}) as a function of the number of self-loops n_s (see Table 8.2).

A similar qualitative behaviour is obtained with the other three configurations (AUGM-OR, CONST-RND, AUGM-RND), has shown respectively in Figure A.1, Figure A.6 and Figure A.11 of the appendix A.

A different perspective is shown in Figure 8.18 where we plotted the empirical cumulative distribution function (ECDF) of each of the 16 differ-

ent attractors number distributions related to the CONST-OR configuration presented in Figure 8.17 with $n_s \in \{0, 1, \dots, 15\}$. Figure 8.18 supports our previous analyses showing that 95% of networks are under the ECDF curve for attractor number gradually bigger. As extreme examples: if $n_s = 0$ then $ECDF \leq 0.95$ for $i \leq 6$, while if $n_s = 15$ then $ECDF \leq 0.95$ for $i \leq 512$.

A generalisation of this pattern can be attained by finding a fit between the ECDF of attractor number and a known discrete distribution. Indeed, for few self-loops the distribution is well fitted by a Poisson distribution. Unfortunately, this fit considerably degrades with increasing number of self-loops; in this latter case, we found a good match with a geometric distribution which tends to reproduce with more accuracy the tail of the attractor number distribution. Nevertheless, the geometric distribution completely misses the peak of the density. Even though continuous, the Weibull distribution [Johnson et al., 1995] provides an overall good trade-off, being defined as a function of two parameters. This result would suggest that the distribution of the number of attractors of the BNs is a mixture of exponential and Rayleigh distribution. A formal study of this issue is planned for future work.

8.4.6 Attractor robustness

The second relevant feature affected by the number of nodes with self-loops is attractor robustness. In our previous preliminary study [Montagna et al., 2018] we observed that the median value of attractor robustness decreases with the number of self-loops. In fact, the distribution of these values is rather wide and a single statistical parameter might miss important features of the phenomenon. Moreover, as the distribution of the number of attractors of the BNs is not uniform, a fair comparison should be achieved by comparing robustness among BNs with the same number of attractors. Therefore, we look here at the minimum and maximum values of robustness— δ_{min} and δ_{max} —averaged across networks with the same number of attractors. In this way we have more balanced results across the 16 different settings, on top of which we can make a comparison and discussion: are self-loops affecting attractor robustness in networks with a different setting but same number of attractors? We computed averages of δ_{min} and δ_{max} on 30 networks randomly selected from the sampled pool of BNs with a given number of attractors; in case 30 networks are not available, statistics are not computed.⁴ We stop computing the averages at the largest number of attractors for which at least 30 sampled BNs have been found.

⁴Indeed, as for some topological configurations and ranges of n_s the computational cost for finding a RBN with a given number of attractors might be extremely high, we considered as statistical significant only averages computed across 30 samples, which is the minimal number suggested by a commonly applied rule of thumb [Cohen, 1995].

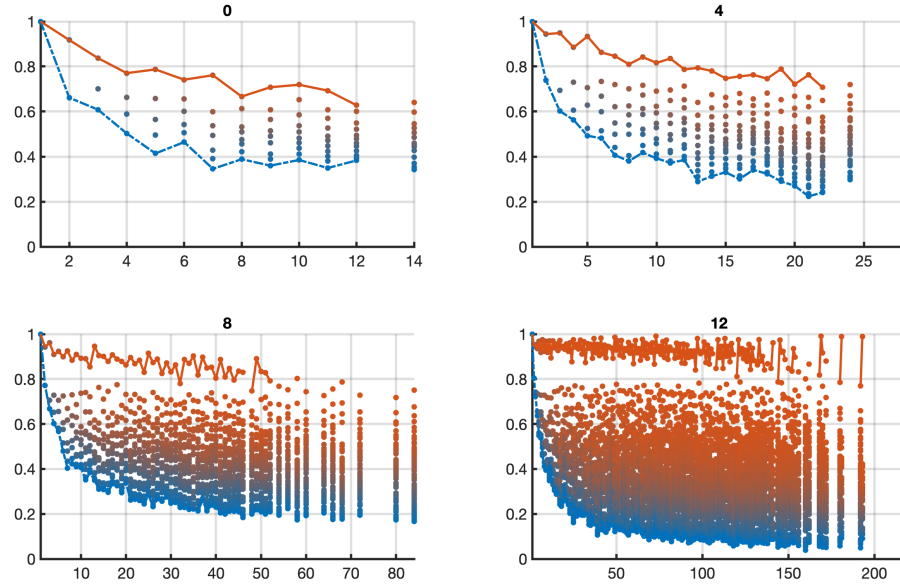


Figure 8.20: δ_{min} , δ_{max} and ATM diagonal values as a function of the number of attractors for CONST-OR configuration. Results are shown for $n_s \in \{0, 4, 8, 12\}$ (see title in each plot).

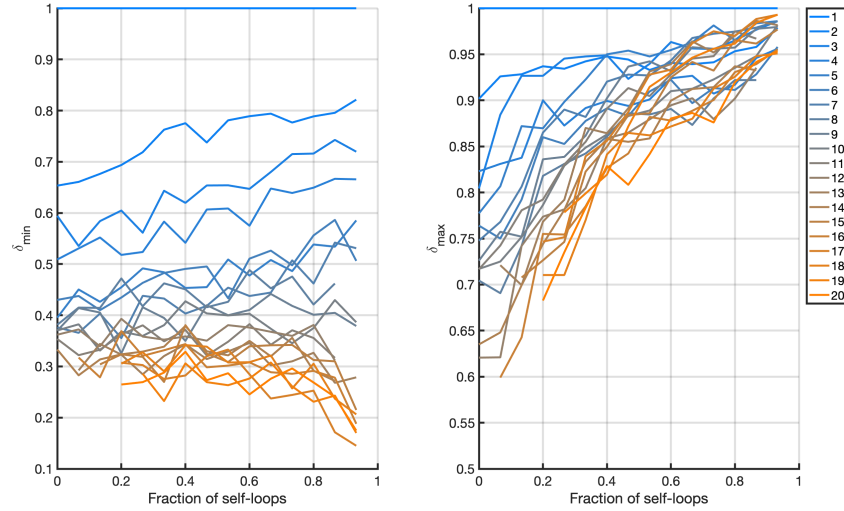


Figure 8.21: δ_{min} (left) and δ_{max} (right) as a function of the fraction of nodes with a self-loop for CONST-OR configuration. Each plot refers to the average values of δ_{min} (left) and δ_{max} (right) on 30 networks with the same number of attractors m , where $m \in \{1, \dots, 20\}$ (see legend).

Figure 8.19 shows the trend of δ_{min} and δ_{max} as a function of the number of attractors across all 16 configurations of self-loops. We can observe that, in each plot, attractor robustness decreases with an increasing number of attractors: both δ_{min} and δ_{max} follow a quasi-monotonically decreasing function (with few local exceptions). This confirms the overall picture we got in our previous work. Results obtained for 15 self-loops has to be interpreted with a bit of ingenuity. In this case, $\delta_{min} = 0$ for all the networks and $\delta_{max} = 1$, independently of the number of attractors. The reason of this phenomenon is that in these peculiar networks in which every node regulates itself in OR, there are at least two fixed point attractors: $(0, 0, \dots, 0)$ and $(1, 1, \dots, 1)$. The first is unstable, as any node flip will switch to 1 at least one node which will keep this value forever; whilst the second is stable for every possible single flip, as any node perturbation $1 \rightarrow 0$ will be immediately reverted at the subsequent update step. In general, we can observe a tendency of δ_{min} to decrease, while δ_{max} increases with the number of self-loops, denoting that variability in attractor robustness increases. This trend is shown in Figure 8.20.

If we restrict the analysis to a limited number of attractors feasible for BNs with any number of self-loops and we consider δ_{max} at a given number of attractors m , we observe a notable fact: the maximum robustness grows with the number of self-loops (see Figure 8.21, right—the trend is shown for $1 \leq m \leq 20$). The picture emerging from the analysis of δ_{min} (Figure 8.21, left) is somewhat more complicated: the trend of minimum robustness is increasing for BNs with few attractors and it decreases in networks with more than 12 attractors, while it does not significantly vary for an intermediate number of attractors. Analogous results, even if less striking, hold for the AUGM-OR. Conversely, the RND cases do not show this behaviour. Figures summarising results for these settings can be found in Figure A.3 (AUGM-OR), Figure A.8 (CONST-RND) and Figure A.13 (AUGM-RND) of the appendix A.

8.5 Discussion and Conclusion

The overall outcome of our analysis is that the addition of self-loops with canalising function to RBNs affects:

- (a) the distribution of the number of attractors, mainly by gradually increasing the maximum number, moving right the peak of the density function and making the distribution flatter; in particular as the fraction of self-loops increases, density function peak becomes gradually less substantial and tails longer;
- (b.1) the **maximum** attractor robustness in two ways: it decreases with the number of attractors, but gradually increases adding self-loops if

compared across BNs with the same number of attractors;

- (b.2) the **minimum** attractor robustness in two ways: it decreases with the number of attractors, and exhibits a composite behaviour if compared across BNs with the same number of attractors but different number of self-loops (it grows for BNs with few attractors and decreases for many attractors, while it is approximately steady for an intermediate number).

Given these experimental results, we claim that self-loops in BNs can positively influence their dynamic behaviour—according to the characteristics required for modelling cell differentiation—but the fraction of self-loops must be accurately chosen to guarantee the best balance between number of attractors (cell types) and attractor robustness (cell type stability). Our claim is that in the range 25–45% of nodes with a self-loop in the network, we can observe a substantial advantage in robustness without exceeding in the amount of attractors. From the analysis we conducted on the Cell Collective [Helikar et al., 2012] database, we found that, among the GRNs with self-loops (54% of the networks available, as discussed in Section 8.4.1), the average value of the fraction of self-loops is 0.2100 with a standard deviation of 0.2104. Details on the distribution are shown in Figure 8.22. This analysis on real networks confirms our hypothesis and leads to the conclusion that a fraction of about 30% self-loops can bring an evolutionary advantage to BN dynamic, especially once modelling cellular differentiation processes.

In particular, simulation results enable us to formulate an evolutionary hypothesis that may be tested *in silico* by means of BNs. Let us suppose that attractors—or sets of attractors—represent cell types. Our conjecture is that autoregulation may have appeared in evolution as a functional component that makes it straightforward to (i) increasing the number of attractors (i.e. cell types) without severely perturbing the other dynamical properties of the network and (ii) consolidating dynamical attractors, e.g. by increasing the robustness of some of them (in other words, to increasing the maximum attractor robustness in a BN with m attractors, a moderate rewiring adding self-loops would be a quite effective procedure). Indeed, a system is evolvable if, subject to mutations on its structure, it exhibits variability in phenotypic traits that may undergo selection. Besides this, by being quite simple and local modifications, self-loops are good candidates as mutation perturbations in evolutionary schemes. As a future work, we plan to investigate this evolutionary hypothesis.

Some questions may be raised concerning the properties of the model we studied in comparison with RBN models studied in the current literature. A first question may arise about the dynamical regime—ordered, disordered or critical—of RBNs with self-loops.⁵ To the best of our knowledge, this

⁵The topic of BNs dynamical regime is rather wide and out of the scope of this work;

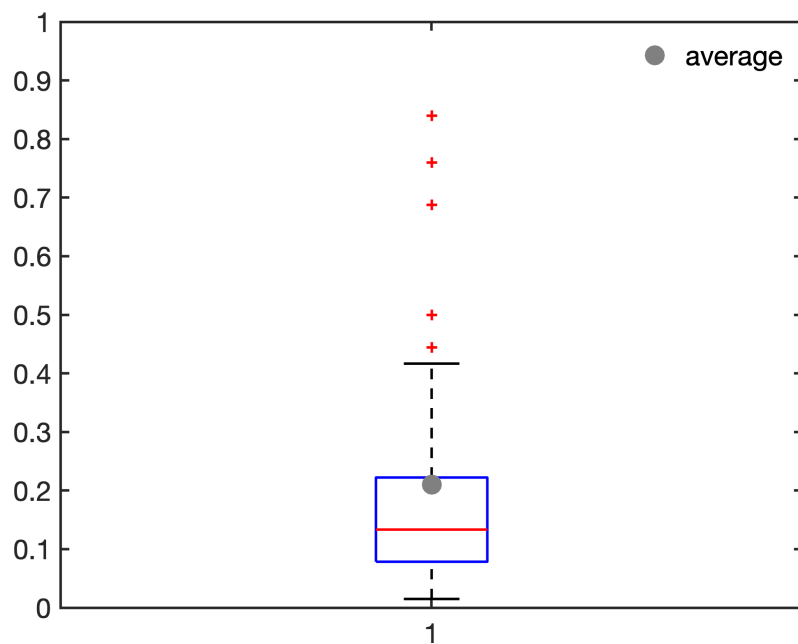


Figure 8.22: Distribution of the fraction of nodes with a self-loop in real networks from the Cell Collective [Helikar et al., 2012] database.

property has not yet been studied so far and we are currently investigating it. However, it should be observed that in the case in which OR functions are introduced with self-loops, the canalising effects are very likely to keep the networks in an ordered regime. Moreover, one may ask to what extent the results obtained for CONST-OR and AUGM-OR differ from those that can be attained in classical RBNs in which the function of some nodes is forced to be an OR. In fact, we addressed this question in a previous work [Montagna et al., 2018] and found that the two models produce strikingly different results. Therefore, the effect of self-loops in RBNs can not solely be ascribed to the specific Boolean function used, but it crucially depends on the topological feature of autoregulation.

we refer the interested reader to specialised literature in dynamical criticality [Roli et al., 2018, Villani et al., 2018, Villani et al., 2017].

Chapter 9

Wrap-up

In this chapter, we briefly report the theoretical and experimental evidences concerning the effects of self-loops in random Boolean networks (RBN) ensembles¹ and we try to draw some general implications deriving from them. In the following, we will refer to *robustness* as the returning probability to the very same attractor after a perturbation.

1. In RBN ensembles, attractors number statistically increases as fraction of nodes having self-loops increases (see Figure 8.5a);
 - (a) this can also be deduced from the Figure 8.17 where probability density estimate of attractors number as self-loops are added to the network is reported: it is clear that the probability of having few attractors decreases while the probability of having a large number of attractors increases.
2. the median of robustness decreases as the number of self-loops increases (see Figure 8.5b);
3. at the same time, also the mean of the difference between MAX and MIN robustness increases with self-loops (see Figure 8.7);
 - (a) this can be observed also in the Figure 8.19;
4. in addition, when we keep the number of attractors fixed the average MAX of robustness increases with self-loops (see Figure 8.21);

Since the *state space* is finite— 2^N possible states for a BN with N nodes—there is a trend of inverse proportionality between basin sizes and the number of attractors, this trend is more evident in RBN ensembles. From points 1, 1a and 2, we can infer that there is a correlation relationship between

¹In particular, this discussion refers to RBN ensembles with self-activating functions, such as logic OR, this is indeed the ensemble presenting the more striking effects that will be showed.

the attraction basins size and the robustness on average ². Based on this hypothesis and the increase of the *diversification parameter* ³—reported in 3 and 3a—we can say that in RBN ensembles an increment in the number of nodes with self-loops leads to a decrease of the MIN value of robustness, on average. The previous statement seems to be true even if the MAX of robustness tends to be larger in networks with a higher number of self-loops—if we compare BNs with the same amount of attractors. Indeed as we can clearly see in Figure 8.19 the MIN negative slope is larger than the positive slope of the MAX.

Therefore, for what concerns the impact of self-loops with self-activation effects (e.g. OR) on ensembles of RBNs we can say that—statistically—some attractors (at least one) tend to be fairly unstable while some others assume stronger attractive capabilities, respectively by lowering or increasing the transition barriers between them.

²Actually there is a correlation between basin sizes and the specific definition of robustness we have adopted.

³See Section 8.2.3 for the definition of the diversification parameter.

Chapter 10

Epigenetics-driven differentiation in BN models

Epigenetic mechanisms play a crucial role on the gene expression regulation (see Section 1.2.3) and—as we have briefly reported in Section 1.3.1—this highly affects the differentiation process itself. For this reason, after introducing the state of the art related to the modelling approaches of epigenetic processes for understanding their role in cell's dynamics, our contributions about the introduction of epigenetics in a Boolean model of differentiation will be presented.

10.1 State of the art

Several mathematical approaches have been proposed with the aim of disentangling the effects of epigenetics in development, differentiation and also in the establishment of aberrant cellular states—like cancer. Noteworthy is the work [Miyamoto et al., 2015] in which the authors investigate the mechanisms of differentiation and cellular reprogramming introducing a continuous model of a minimal gene regulatory network (GRN) able to give rise to both pluripotent and differentiated states. In their modelling approach, an epigenetic process—introduced as a gene expression fixation—turns out to be important to increase the stability of the attained differentiated states and to reproduce with more accuracy the phenomenology of the reprogramming process.

In the works [Turner et al., 2017, Turner et al., 2013], the authors have ascertained that the addition of an epigenetic layer—in the form of Boolean switches that dynamically change the actual network topology—within recurrent neural networks lead to better performance in the achievement of certain target tasks, as compared to models without it.

To the best of our knowledge, the specific role of epigenetics in the dynamics of discrete models of GRN has been addressed only by [Bull, 2014].

The author does not focus on the differentiation process as such, but instead, he evaluates the potential of Random Boolean networks (RBNs) with epigenetic control—which is interpreted as additional nodes that change the regular transcription dynamics—in NK landscapes [Kauffman and Levin, 1987].

10.2 A simplified model of chromatin dynamics drives differentiation process in Boolean models of GRN

As we already know, eukaryotic cells are characterised by the organisation of DNA in a condensed structure, called chromatin. Methylation, a well studied epigenetic mechanism, influence the gene regulation process by acting on the compactness of the chromatin structure. It most often leads to tightly packed regions of chromatin called **heterochromatin** [Gilbert and Barresi, 2016, Perino and Veenstra, 2016, Schuettengruber and Cavalli, 2009]. These regions are not accessible neither by transcription factors nor by RNA polymerases and so the expression of genes belonging to these DNA areas is inhibited. Biological cells exploit differential methylation to modulate their gene expression during development and differentiation. It is important to note that, along lineages, the attained configurations of DNA methylation are inherited and progressively extended as cells become more specialised [Kim and Costello, 2017]. Therefore, methylation contributes to maintain and stabilise the attained gene expressions that ultimately characterise the identities of the various cell states.

Kauffman and Huang [Kauffman, 1969b, Huang et al., 2009b, Huang et al., 2005, Huang and Ingber, 2000] have laid the foundations for a rigorous mathematical description of GRN dynamics in terms of dynamical systems, with attractors that model cell types. Serra and Villani [Villani et al., 2011, Villani and Serra, 2013a, Serra et al., 2010], subsequently, have raised the abstraction level bringing the attention to the relations among attractors—and set of attractors—under noise influence, describing the cell specialisation process by a progressive decrease in noise. The detailed differentiation properties that the model is able to reproduce are reported in Section 6.5.

The model focuses on the dynamics of a single cell represented as an *autonomous* system¹ subject to intracellular noise. Cell types are defined as the portions of the space of states in which the dynamics remains trapped, under a specific noise level. Changes in the intracellular level of noise drive the differentiation process: high noise levels correspond to pluripotent cells

¹Here we adopt the terminology of dynamical systems in which the adjective *autonomous* is used to denote systems that are not subject to inputs, therefore their state may change in time only owing to internal mechanisms.

while the low levels to fully differentiated ones. Experimental analysis on RBNs subject to stochastic dynamics [Braccini et al., 2018] and the successful evolution of networks able to attain not trivial differentiation dynamics [Braccini et al., 2017, Benedettini et al., 2014] proved the expressiveness and plausibility of this model.

Differentiation represents a major challenge for every model of gene regulatory networks that, like RBNs, is based on deterministic dynamical systems which asymptotically reach stable attractor states, to be identified with the different cell types. Indeed, under the action of the deterministic dynamics, a stable attractor does not change any longer so it must represent a fully differentiated cell type. Therefore cells which are found at intermediate differentiation levels (e.g. pluripotent cells) should be associated to transients—an unsatisfactory proposal, since it is known that there exist long-lived pluripotent cells, which should rather be represented by metastable states.

The way out of this conundrum requires a mechanism to escape from the deterministic attractors. While this mechanism is provided in our previously described model by means of intracellular noise, in this work we want to explore an alternative—complementary—possibility, i.e. that it is due to an external signal. In this way, the system is no longer autonomous, and escaping from the attractors of the corresponding deterministic system becomes possible. External signals are indeed known to affect embryo evolution, and the simplest way to describe their effect in a GRN model is that of clamping the values of some network nodes to fixed values.

10.2.1 The model

As previously discussed, methylation—even if it is not the only phenomenon in place—has a non negligible impact on cell fate determination and maintenance. Here we are especially interested in its abstract role in simplified models of GRNs, namely in Boolean networks. Indeed, borrowing the idea of a progressive methylation state of the chromatin along the development and differentiation of biological cells, we propose an analogous mechanism in BN models. Similarly to what happens in the heterochromatin condition, the expression of some BN nodes is blocked to value 0; these nodes will be referred to as **frozen** in the following.

Theoretically, the formulation of this peculiar methylation mechanism implies a sort of simplification of the network, as it reduces the nodes that are actually subject to a dynamic update, and so restricting the number of combinations that the system itself can assume. Therefore, it is not a priori clear whether this mechanism can accommodate **path dependent differentiation**: cell types determined by the specific sequences of methylated genes.

This model relies on the hypothesis—to be verified in RBNs—that the progression of frozen nodes imposes the arrow of time of the differentiation

process and, at the same time, different patterns of methylated nodes give rise to distinct lineages, and so cell types. Indeed, biological differentiation is characterised by the presence of different stages of differentiation and by progressively specialisation of cells.

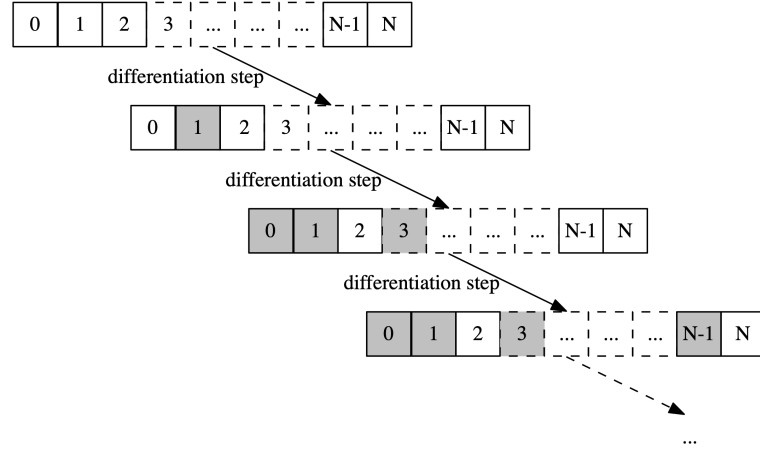


Figure 10.1: Schematic representation of the methylation mechanism introduced. Grey nodes represent frozen nodes (nodes constrained to assume the value 0, regardless of the actual values of its inputs). The specific patterns of frozen nodes in the ensembles that represent the state of the BN over time have no other meaning than to exemplify the methylation process introduced.

A schematic representation of this *Boolean methylation-inspired mechanism* is depicted in Figure 10.1. In this work we undertake an experimental analysis of the main dynamical properties of RBNs subject to this process of progressive methylation. For this mechanism to be useful in a plausible BN differentiation model, it should (i) progressively stabilise the network and (ii) give origin to different lineages depending on the nodes chosen to be frozen. If these properties are attained in RBN ensembles, then we could suppose that evolution may act to tune the dynamics of the network so as to achieve a specific differentiation lineage tree. The choice of setting to 0 the nodes to be frozen is motivated by the inhibition effect of most methylation mechanisms and introduces an asymmetry in the RBNs model, as it progressively bias the Boolean functions to 0. However, this is not a limitation of the model, which can be extended to take into account also actions in which nodes are clamped to 1 and so provide even more variability in the lineages.

10.2.2 Results

The random Boolean networks used in these experiments are subject to a synchronous and deterministic dynamics, therefore fixed points and cycles are possible asymptotic states. For all the experiments, statistics are taken across 100 RBN with $n = 500$ and $k = 2$. We focused only on networks with $k = 2$ because the size of the network, combined with the other chosen parameters, would have made the experimental analysis computationally prohibitive. The Boolean functions are defined on the basis of the bias parameter p , which defines the probability to assign value 1 in a row of a node truth table. The variation of the parameter p makes it possible to determine the dynamical regime of the system (ordered, critical or chaotic) [Bastolla and Parisi, 1997]: so, the limitation due to the choice of a specific connectivity is thus eliminated. Since we want to analyse the emerging generic properties induced only by the proposed methylation mechanism in ensembles of RBNs, we used an *exact* bias. Exact bias is computed by generating each time a random permutation of a vector of Boolean values with a length equal to the number of nodes in the network and a fraction p of 1's, and by using partitions of this vector to define the output values of Boolean functions. In this way, we remove from the statistics any possible contribution produced by a variance in network dynamic regime. We generated RBNs with $p = 0.1$, i.e. in the ordered regime, and $p = 0.5$, corresponding to the critical regime. As results with ordered RBNs are rather uninformative, we only show results for critical RBNs.

Attractor number distribution To providing the trend of the number of attractors as the fraction of frozen nodes increases we generated 100 RBNs and for each number of frozen nodes we performed a search of the attractors starting from 10^4 random initial states. The range of frozen nodes considered varies from 0 to 200 with a step of 5 nodes. Boxplots showing the distribution of the number of attractors as a function of the number of frozen nodes are depicted in Figure 10.2, along with the mean of these distributions. As expected, the number of attractors decreases with the number of frozen nodes, even though it remains non negligible up to one fifth of frozen nodes. A question may arise as to how many attractors are fixed points, as one expects an increasing number of fixed points as the RBNs become more ordered. This expectation is indeed confirmed, as shown in Figure 10.3.

Derrida parameter With the aim of assessing the intuition suggesting a progressive shift towards an ordered regime of the ensemble of RBNs subject to the methylation mechanism, we computed the distribution of the Derrida parameter [Bastolla and Parisi, 1997] λ , computed after one step. This parameter is evaluated by taking, for each state considered (10^3 in total), the means of the Hamming distances after one update between the state and

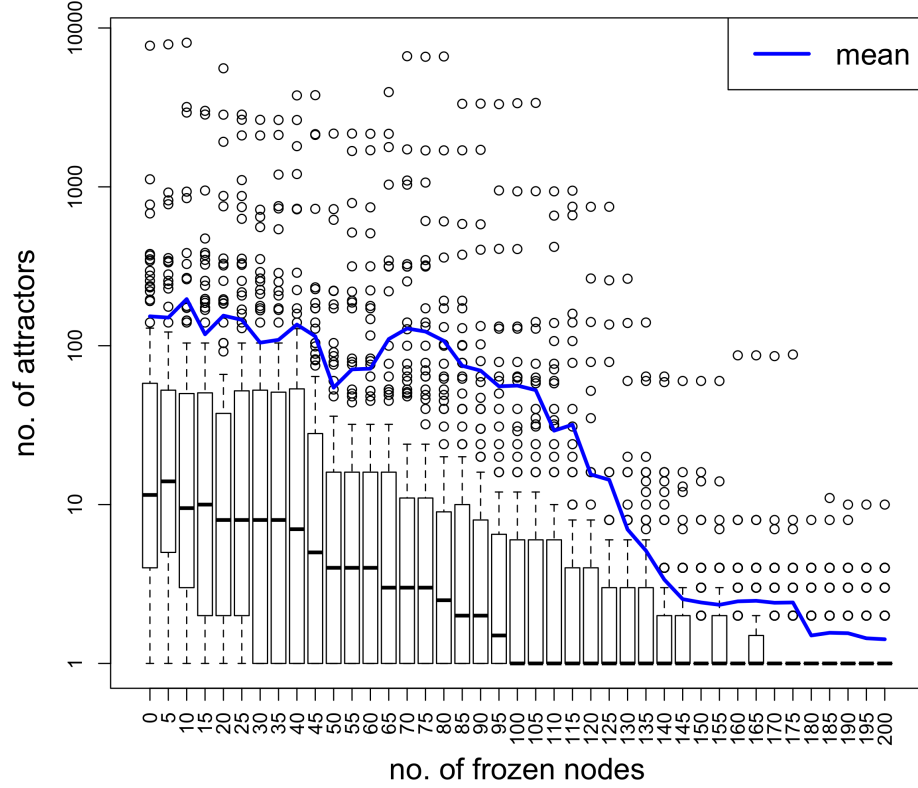


Figure 10.2: Distribution of the number of attractors for the configuration $n = 500$, $k = 2$, $p = 0.5$ as the number of frozen nodes increases from 0 to 200 with a frozen step of 5 nodes at a time. The continuous line illustrates the trend of the mean.

the perturbed one (a logic negation of a single node value) in all not frozen nodes, taken one at a time. In particular, statistics report the distributions of the 100 means of the means, one parameter value for each RBN which summarises the overall behaviour observed along the 10^3 random states. For this investigation, we consider a number of frozen nodes represented by a percentage of $\{0, 10, 20, 50\}$ of all nodes. Figure 10.4 depicts the boxplots summarising the distribution of λ for the ensembles sampled; the trend towards an increasing order is confirmed (the results for $p = 0.1$, on the left in Figure 10.4, are provided as a comparison)².

²According to [Chambers et al., 1983], although it does not represent a formal test, if the notches of two boxplots do not overlap, the difference between the medians of the relative distributions can be considered significant, or there is “strong evidence” (95% confidence) that their medians differ. Since the formula for calculating the notch value is the following $median \pm 1.57 \times IQR/n^{0.5}$, we can conclude that when the number of

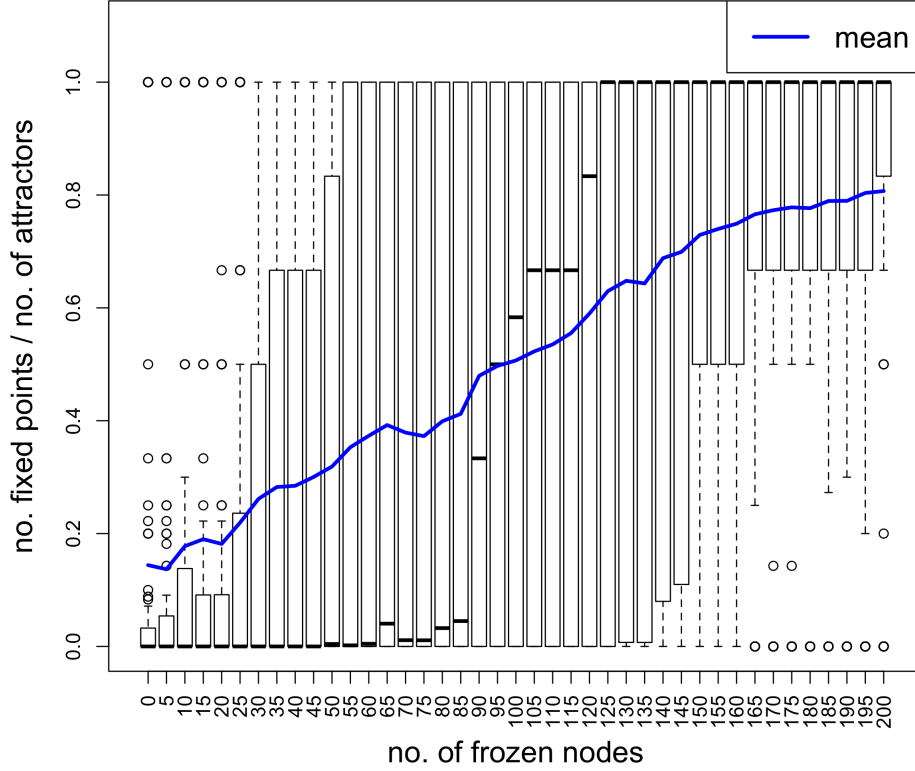


Figure 10.3: Distribution of the number of fixed points over the number of attractors for the configuration $n = 500$, $k = 2$, $p = 0.5$ as the number of frozen nodes increases from 0 to 200 with a frozen step of 5 nodes at a time. The continuous line illustrates the trend of the mean.

The results shown so far support the conjecture that a progressive freezing pushes RBNs towards order. One may argue that a result not in agreement with this expectation might indeed sound surprising, nevertheless it is important to assess it experimentally in particular because this trend is not trivial at all in finite-size RBNs. While in infinite-size RBNs just a tiny fraction of frozen nodes leads to a complete stasis of the network³, in finite-size RBNs we observe that the number of attractors and the Derrida parameter are kept at significant values even in the presence of a non-negligible fraction

samples n is greater than or equal to four, it remains inside the *IQR*, and therefore inside the rectangle. Having said that and observing that there is no overlap between the boxplot rectangles we can conclude that there is statistical evidence in favour of rejecting the null hypothesis: the differences between the medians are not very likely due to chance.

³A formal model of this behaviour is subject of ongoing work.

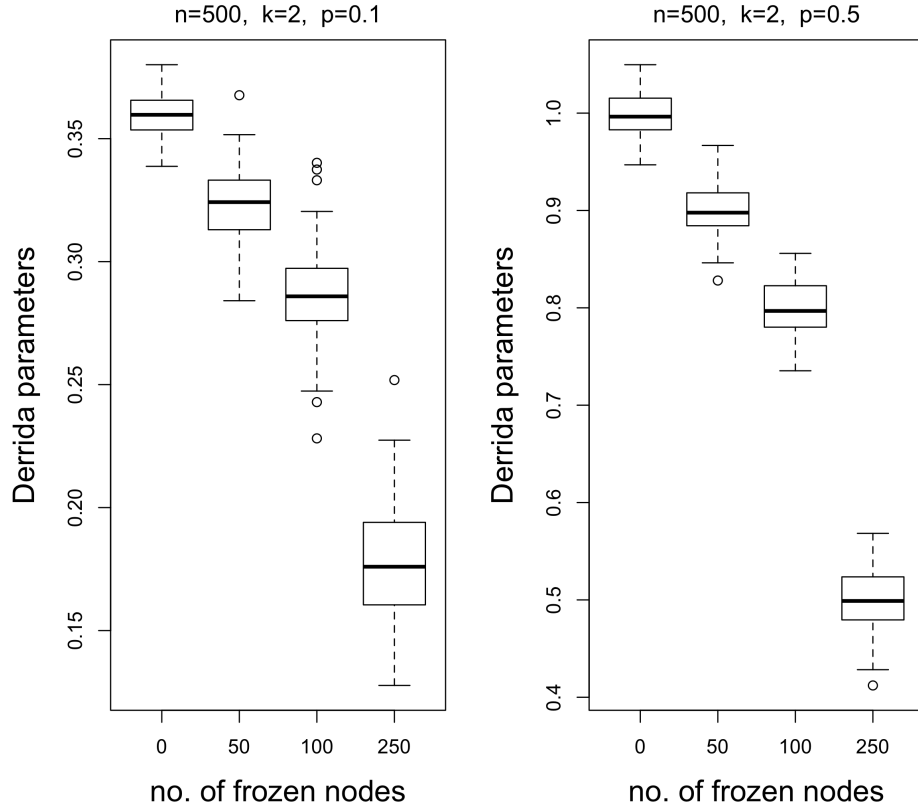


Figure 10.4: Distribution of the Derrida parameters as the number of frozen nodes increases, for both ordered and critical ensembles of 100 RBNs with $N = 500$.

of frozen nodes. This result suggests that in finite-size RBNs, while a progressive freezing tends to increase order in network dynamics, it may still be open to variability. This last characteristic is relevant especially with respect to the possible paths across attractors that are feasible as the consequence of different choices in the nodes to be frozen.

Diversity estimation In previous sections we have summarised with *path dependent differentiation* the property of generating different cell types as a result of different sequences of methylated genes. We can characterise the tendency of this mechanism to give rise to this property by inspecting the *diversity* caused by different combinations of methylated genes at any attained differentiation stage. For this purpose, we generate for each state of the methylation process (state represented by the already frozen nodes

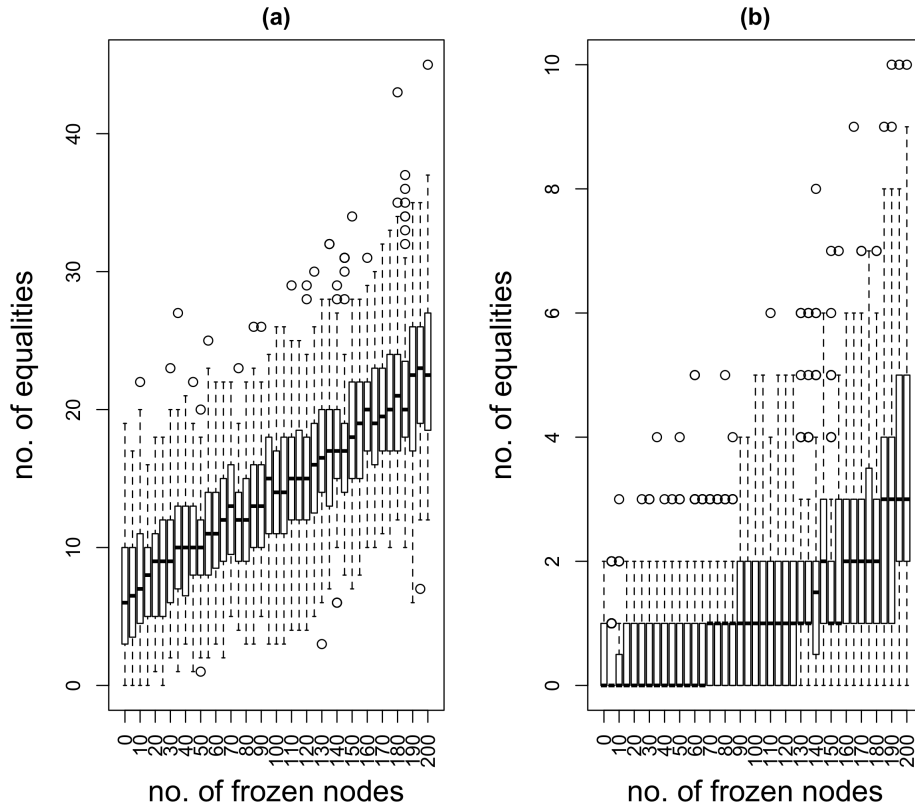


Figure 10.5: The trend of the number of equal reached asymptotic states considered in pairs and after removing the part of the already frozen nodes (x-values) and the set of nodes that constitute the triplets. **(a)** Triplets randomly chosen among all the non-frozen nodes, **(b)** triplets randomly chosen among the non-frozen nodes with value 1.

and the attractor reached) 10^2 couples of triplets of nodes,⁴ among the non-already methylated nodes. This triplet is frozen while the network is in an asymptotic state, therefore after this perturbation the BN dynamics is subject to a transient and subsequently the network can either return to an attractor equal to the current one—except for the frozen triplet—or reach a different one. The freezing step may be taken at any state—i.e. *phase*—of the current attractor; as the phase of the attractor may be a source of variability and here we want to assess the contribution of the choice of frozen

⁴The choice of 3 nodes is somehow arbitrary, but motivated by the requirement of involving a small number of nodes to be frozen, while keeping the possibility of significantly perturbing the attractor. However, previous preliminary experiments on different network size and number of frozen nodes confirm the qualitative behaviour we show in this work.

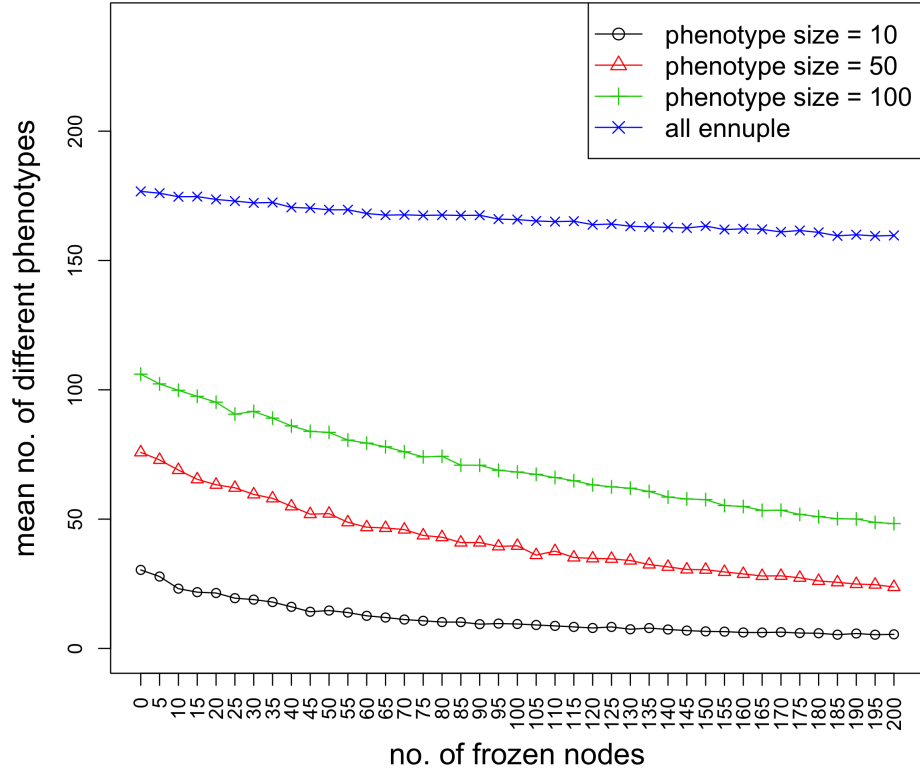


Figure 10.6: The trend of the number of *diversities* caused by 200 triplets of frozen nodes. **The triplets are randomly chosen among the non-already frozen nodes.** Diversities are measured by considering if the reached attractors are different (all tuples case) or by means of randomly chosen patterns (of sizes equal to 10, 50, 100) which select the nodes on which perform the comparison between the reached attractors.

nodes only, once the attractor is reached after freezing a triplet of nodes, its minimum state according to the lexicographic order is chosen. As networks are random, this choice does not introduce any bias and in this way we rule out any possible contribution of attractor phase in the diversity of paths originated by freezing steps.⁵ The diversity is then measured depending on the characteristics of the new asymptotic states on which the dynamics settles after the triplet is frozen. As we aim at providing general results, not bound to a specific definition of *phenotype*⁶, which should be supported by motivations on a concrete biological case, we analyse the arising diversities

⁵In other words, we pose us in the condition of minimal diversity.

⁶See the following parts of the text for a more detailed discussion.

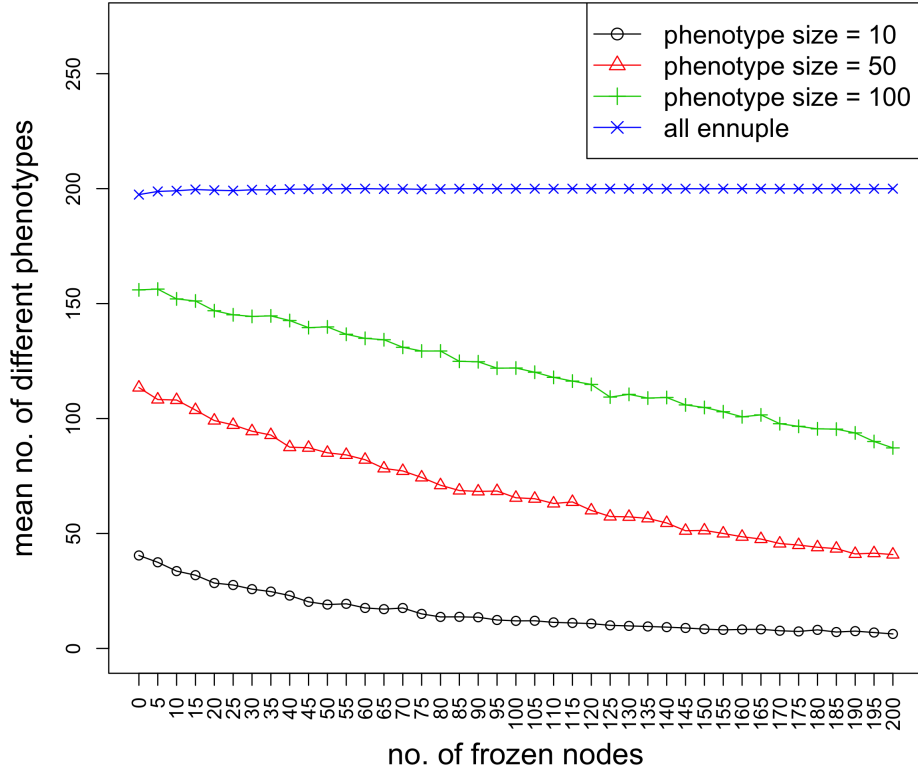


Figure 10.7: The trend of the number of *diversities* caused by 200 triplets of frozen nodes. **The triplets are randomly chosen among nodes that have value 1, in the chosen phase of the starting attractor..** Diversities are measured by considering if the reached attractors are different (all tuples case) or by means of randomly chosen patterns (of sizes equal to 10, 50, 100) which select the nodes on which perform the comparison between the reached attractors.

in various condition. Particularly, we count:

- the number of equal reached asymptotic states considered in pairs and after removing the part of the already frozen nodes and the set of nodes that constitute the triplets;
- the differences among all the reached attractors caused by the generated triplets, by considering subset of genes (patterns in the following) of different sizes (10, 50, 100) randomly chosen;
- the differences among all the reached attractors caused by the generated triplets, by considering the states vectors in their entirety.

By doing so we will have an overall picture of how this mechanism behaves in ensembles of RBNs, without limiting ourselves to particular points of view. As for the attractors distribution analysis, the range of frozen nodes considered varies from 0 to 200 with a step of 5 nodes. We stress that in this model the various degrees of differentiation are characterised by a distinct number of frozen nodes: the higher the number of frozen nodes the more differentiated the cell types. The triplets to be frozen are chosen at random among all the non already frozen nodes; we also made experiments with conditioning this choice to nodes that assume value 1 in the attractor state chosen for the perturbation. In this way, we can assess the highest level of variability that can be attained, as all the three nodes are actually perturbed by freezing.

The distribution of the frequency of equal pairs of attractors is shown in Figure 10.5; we observe that the median frequency of equal pairs increases from about 7/100 to 20/100 with the number of frozen nodes, while it is limited to low percentages when frozen nodes are chosen among the active ones (value 1). This result shows that the probability of choosing two different triplets⁷ leading to the same asymptotic state after being frozen is rather low; therefore, at least for RBNs with at most 2/5 of frozen nodes, the different paths generated by freezing are a significant fraction of all the possible ones, despite the tendency towards a more ordered regime. This observation is confirmed also by the statistics involving the total number of different patterns. With the term *pattern* we refer to a projection of the network dynamics in subset of nodes. So, patterns in this context define the observable phenotypes in a way strongly related to the concept of macrostate introduced in [Borriello et al., 2018, Moris et al., 2016]. These latter results are shown in Figures 10.6 and 10.7. It is worth observing that, even when differences are estimated on the basis of 10 nodes, the fraction of overall different patterns is still non-negligible up to 100 frozen nodes out of 500.

These results support the hypothesis that different freezing patterns in RBNs are very likely to produce different trajectories along attractors, and therefore variability in differentiation paths can be attained also by means of this mechanism.

10.2.3 Conclusion

We have explored the possibility of incorporating epigenetic mechanisms—methylation in particular—into BN models of GRNs. We focused on those processes responsible for high chromatin compaction, that influences gene transcription by controlling the accessibility of DNA to transcription factors and RNA polymerases. Accordingly, in our model we progressively freeze—i.e. clamp to 0—a subset of nodes and analyse the impact of this modification

⁷As triplets are chosen at random among at least 300 nodes, the fraction of equal ones is negligible.

on network dynamical features, namely on attractor number—in analogy with the number of cell types—, on the Derrida parameter—to assess the extent to which RBNs with frozen nodes tend to an ordered regime—and on attractor diversity as attained by different combinations of frozen nodes.

We observed that the number of attractors in RBNs decreases with the number of frozen nodes and the same does the Derrida parameter, suggesting that, from an ensemble point of view, the larger the fraction of frozen nodes the more ordered the RBNs. These results are in agreement with the intuition that, by clamping to 0 a fraction of RBN nodes, not only the state space is reduced with respect to the original network, but frozen nodes absorb perturbations and so they favour network stability. These properties are to some extent the abstract counterpart of progressive reduced alternatives and stability along differentiation stages. Moreover, results show a very interesting property of RBNs: they maintain diversity in terms of possible asymptotic states originating from different combinations of frozen nodes, both during the process of progressive freezing itself and in the final reached states. We assessed this diversity by means of three metrics, so as to attain general results. We found that different choices in nodes to be frozen are very likely to lead to different asymptotic states, implying that diverse differentiation paths can be generated. As expected, this diversity tends to decrease with the fraction of frozen nodes in the network.

10.2.4 Future work

As future work, we plan to add in our model mechanisms to reproducing open chromatin structure, where genes are made more accessible and their transcriptions eased. The combined effects of both closing and opening chromatin structure on attractors and other relevant features of BNs will be consequently analysed. Moreover, since epigenetic is expected to have an impact on cell type stability, we are devising a set of experiments to measure how attractor robustness changes along the path of differentiation, for example by measuring the impact of external signals—possibly modulated—during different stages of differentiation. To conclude, epigenetic is only one of the factors that are responsible for cell type transitions and definitions. Signalling cues, typically generated by other cells, are another crucial actor in the process of differentiation. In this perspective, we are planning to study models involving networks of BNs, so as to explore the possibility of modelling differentiation in a multi-cellular setting.

10.3 The effects of a simplified model of chromatin dynamics on attractors robustness in RBNs with self-loops

The methylation mechanism that we have proposed statistically produces—on random Boolean networks (RBNs) ensembles—a decrease of the attractor number and a dynamics that tends to behaviours resembling ordered RBNs. However, this mechanism does not preclude the possibility of generating **path dependent differentiation**, i.e. cell types determined by the specific sequence of methylated genes.

In [Braccini et al., 2019b], we analyse the effects of the number of methylated (*frozen*) nodes on attractors robustness, defined as the probability of returning to the same attractor after a temporary node perturbation. Furthermore, since self-loops perform important functions in processes of real biological GRNs [Raj et al., 2010] and given that lately some works have begun to shed light on their role in Boolean model of GRN [Montagna et al., 2018, Montagna et al., 2020]—especially related to differentiation phenomenon [Braccini et al., 2019a]—we investigate the change of attractors robustness as a function of the combined effects of the methylation process and the presence of activating self-loops.

The motivation supporting this study is to extend classical RBN models so as to capture differentiation phenomena more accurately and provide a model suitable for comparisons with real data. On the one hand, the methylation mechanism based on clamping nodes at 0 has the main effect of stabilising the network; on the other hand, self-loops may increase the number of attractors in a RBN⁸ and may reduce the average probability of returning to the same attractor after a perturbation. Therefore, the main question is whether these two mechanisms tend to compensate each other and a balanced combination of the individual effects is attained and, if so, under which conditions this happens.

10.3.1 Experimental settings

For all the experiments, statistics are taken across 100 RBNs with $n \in \{20, 50\}$, $k = 2$ and $p = 0.5$. The BNs used in these experiments are subject to a synchronous and deterministic dynamics. In BN models, cell types may be represented by attractors—or sets of attractors, depending on the interpretation chosen. We measure the *robustness* of an attractor as the probability of returning to it after a temporary flip of the value of a randomly chosen node, also called *robust adaptation*. In our experiments we sampled the possible transitions between attractors after a perturbation and

⁸This happens in particular when the self-regulation is modelled by a canalising function, such as the logical OR.

Table 10.1: Summary of the experimental parameters.

Number of nodes (n)	Number of self-loops (N_{sl})	Number of frozen nodes (N_f)
20	0, 3, 6	0, 3, 6, 9
50	0, 5, 10	0, 5, 10, 20

recorded the *returning probability* to each of the attractors sampled.

For each network we took the average returning probability computed across all the attractors—omitting the cases with only one attractor. Experiments were performed for a varying number of frozen nodes and number of self-loops N_{sl} , depending on the size of the network (see Table 10.1). The number of self-loops considered for $n = 50$ had to be limited due to the exponential increase of attractors number, which might make the computation of the attractor returning probability computationally impractical. According to previous experimental settings [Montagna et al., 2020], nodes with self-loops are chosen at random and ruled by *OR*, *AND* or random (RND) logical functions (i.e. $p = 0.5$).

10.3.2 Results

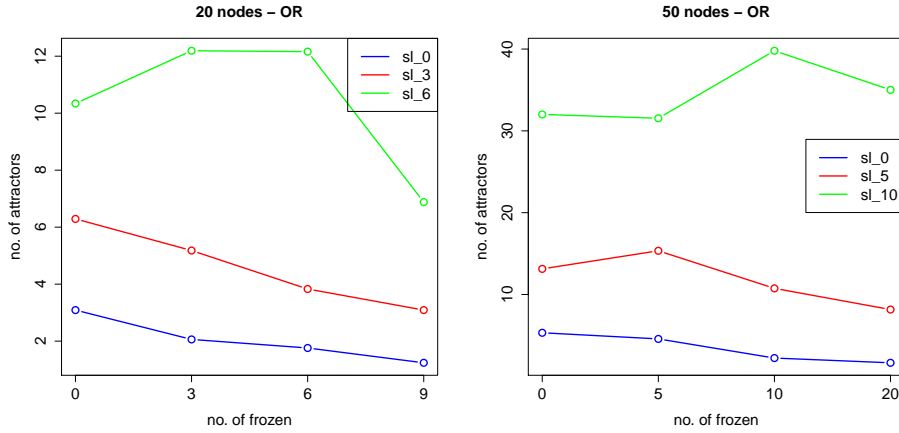


Figure 10.8: Mean attractor number in BNs with 20 (left) and 50 (right) nodes for several combinations of number of self-loops and frozen nodes. Self-loops added in OR.

The first subject of our analysis is the number of attractors. Indeed, attractors play a fundamental role in BN models of differentiation, as attractors—or sets of thereof—are associated to cell types. Moreover, the number of attractors of a BNs is a reckoning of the possible equilibrium states the system has and so the possible ‘answers’ it provides to perturbations: as stated by Kauffman [Kauffman, 2000], a real cell should be able to discriminate

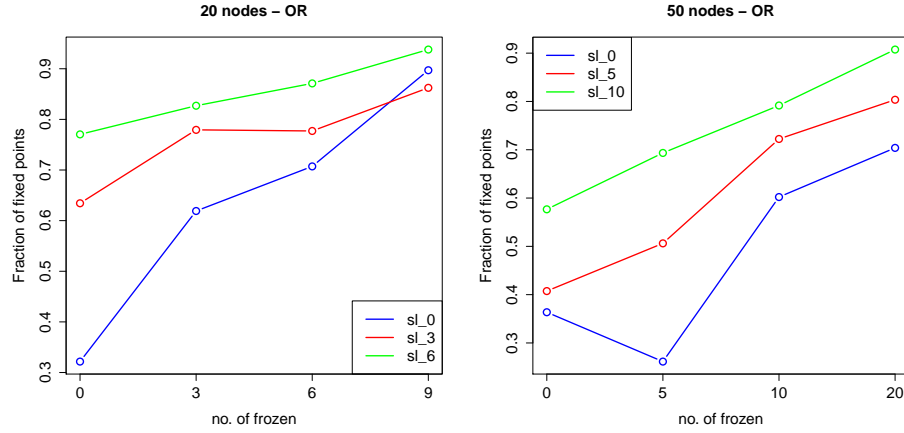


Figure 10.9: Mean fraction of fixed points in BNs with 20 (left) and 50 (right) nodes for several combinations of number of self-loops and frozen nodes. Self-loops added in OR.

among a number of classes that provides the optimal balance between classification of the environmental stimuli and reliable and robust classification. Figure 10.8 shows the trend of the average number of attractors in the case of self-loops in OR. We observe that for a fixed number of frozen nodes, the number of attractors increases with self-loops. This result confirms previous results on RBNs with self-loops. The trend at fixed number of self-loops is instead quite informative: frozen nodes tends to compensate the increase of the attractors number. This result is striking in the case of 20 nodes RBNs, whilst for $n = 50$ the effect is less marked even though the containment of the number of attractors is anyway clear.

It has also been observed that the fraction of fixed points among the attractors increases with the number of frozen nodes or the number of self-loops, see sections 8.4 and 10.2. Figure 10.9 shows the fraction of fixed points, averaged across all the attractors. As expected, the net effect is that fixed points increase both with the number of frozen nodes and the number of self-loops.

These results on the number of attractors are obtained in the case of self-loops in OR, which have the property of keeping indefinitely the value of a node at 1 after it has reached this value during the regular updates of the network. On the other side, freezing a node means setting it to 0 forever. One may ask what is the effect of using different Boolean functions in the nodes with self-loop, in particular in the case of AND functions which, in a sense, impose the same bias as freezing towards zero. Results of this latter case are summarised in Figures 10.10 and 10.11, where we can observe that the combined effect is to drastically reduce the number of attractors. A similar trend, although somehow diluted, is obtained when self-loops are

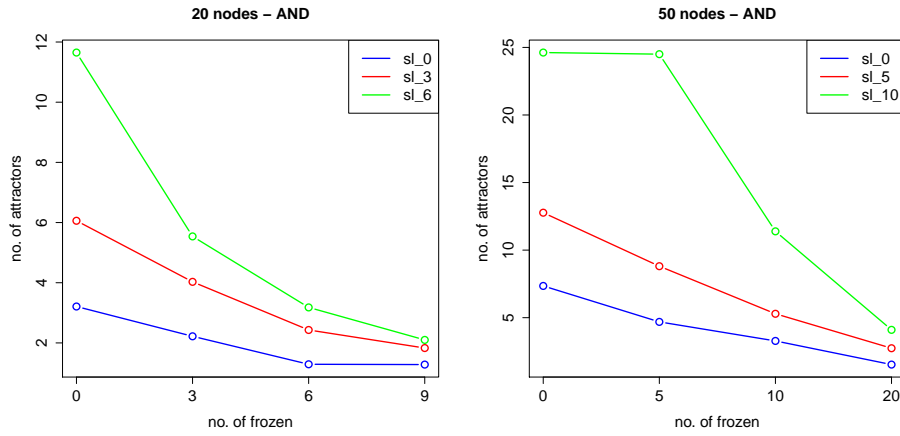


Figure 10.10: Mean attractor number in BNs with 20 (left) and 50 (right) nodes for several combinations of number of self-loops and frozen nodes. Self-loops added in AND.

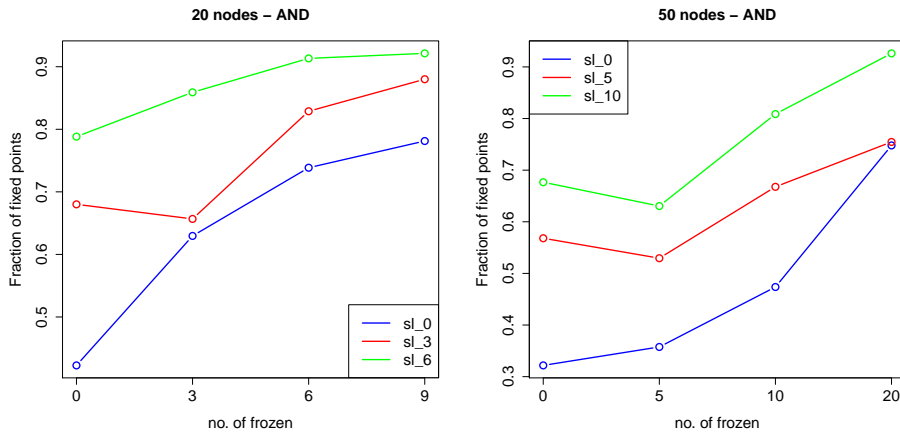


Figure 10.11: Mean fraction of fixed points in BNs with 20 (left) and 50 (right) nodes for several combinations of number of self-loops and frozen nodes. Self-loops added in AND.

associated to random functions (see Figures 10.12 and 10.13).

Results on average attractor robustness are shown in Figure 10.14, in the case of self-loops in OR, where the mean robustness is plotted against N_f for all the values of N_{sl} . If we focus on the trend of a single curve, we observe that the average robustness monotonically increases with the number of frozen nodes. This result is probably not surprising but it is the first time it is experimentally assessed. When we consider the trend as a function of the number of nodes with self-loops we observe a non-monotonic behaviour:

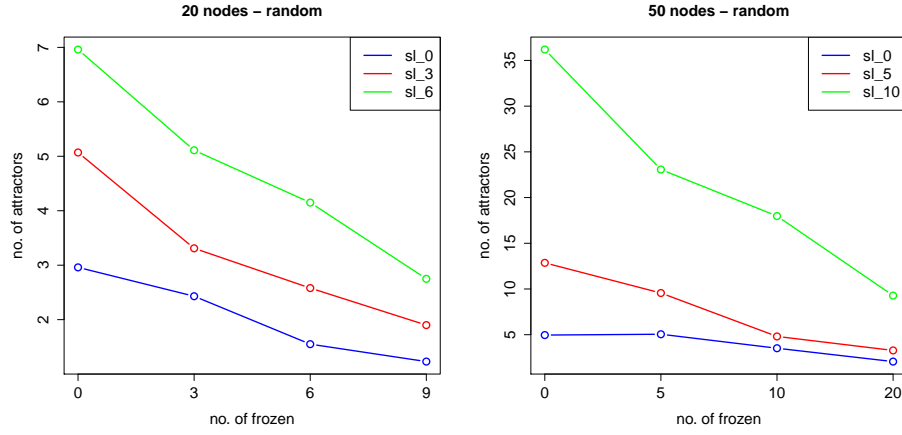


Figure 10.12: Mean attractor number in BNs with 20 (left) and 50 (right) nodes for several combinations of number of self-loops and frozen nodes. Self-loops added with a random function.

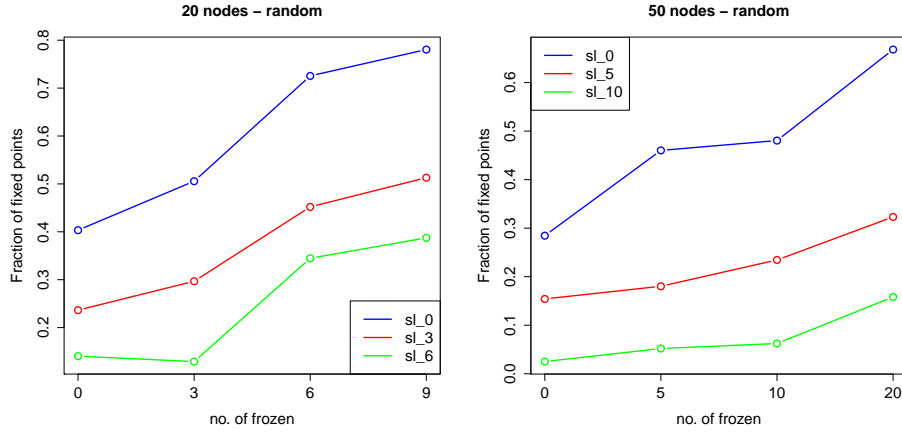


Figure 10.13: Mean fraction of fixed points in BNs with 20 (left) and 50 (right) nodes for several combinations of number of self-loops and frozen nodes. Self-loops added with a random function.

a moderate number of self-loops is further reinforcing the effect of frozen nodes, but a greater one has detrimental effects on robustness. This result is in agreement with previous ones [Montagna et al., 2018], where it was shown that a fraction of nodes with self-loops higher than about 20% makes average robustness drop. These results support the hypothesis that a mild fraction of nodes with self-loops is beneficial for RBNs attractor robustness, especially if combined with freezing mechanisms that model methylation. Current biological cells are the result of evolution, therefore Boolean models of them

are not random; nevertheless, if random BNs with our variants are proven to exhibit features closer to the ones of real cells than simple RBNs, then this enriched RBN model is likely to be a more accurate model of real cells and capture relevant phenomena with higher accuracy. Moreover, in an evolutionary perspective, RBNs with self-loops and methylation mechanisms provide a more suitable starting condition for the evolution of models towards given target characteristics. The case of self-loops with AND function provides a somewhat different picture, because the tension between frozen nodes and self-loops is no longer present and the average robustness tends to increase (see Figure 10.15). Analogous considerations hold for self-loops with random functions, even though less marked (see Figure 10.16).

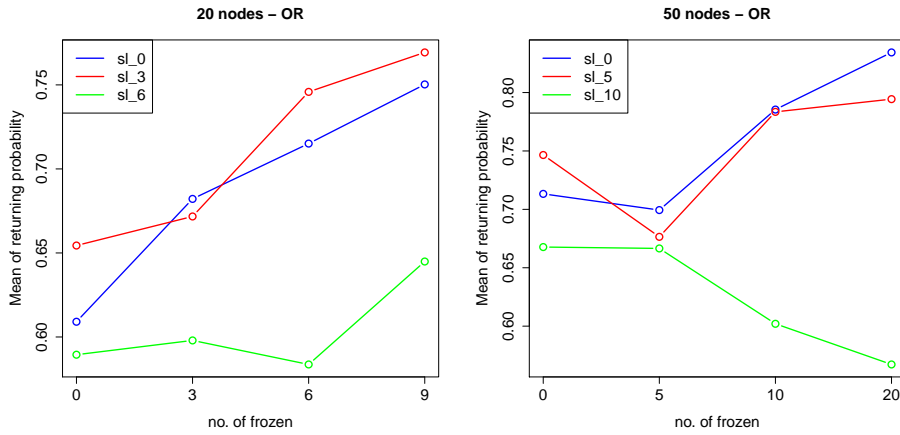


Figure 10.14: Mean attractor robustness in BNs with 20 (left) and 50 (right) nodes for several combinations of number of self-loops and frozen nodes. Self-loops added in OR.

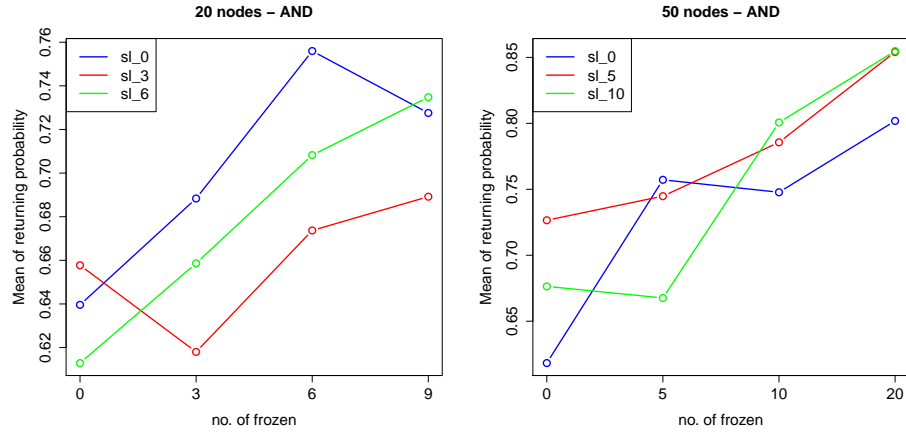


Figure 10.15: Mean attractor robustness in BNs with 20 (left) and 50 (right) nodes for several combinations of number of self-loops and frozen nodes. Self-loops added in AND.

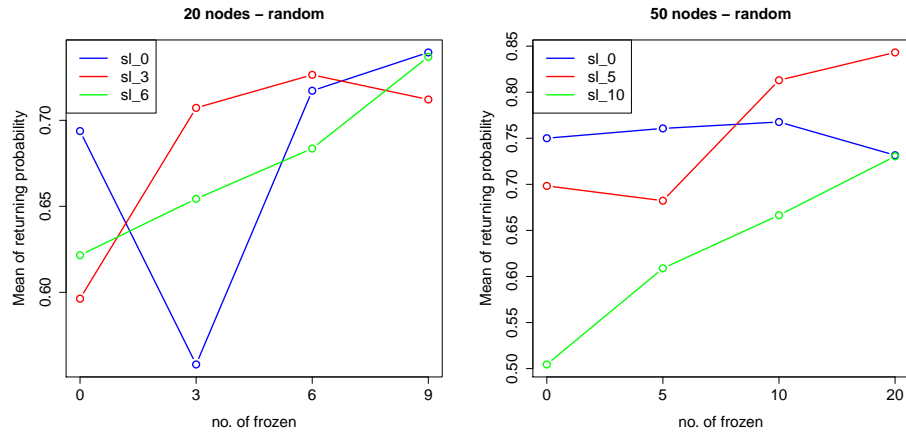


Figure 10.16: Mean attractor robustness in BNs with 20 (left) and 50 (right) nodes for several combinations of number of self-loops and frozen nodes. Self-loops added with a random function.

Part IV

Differentiation models in Robotics

Chapter 11

Dynamical Systems Robotics

As we have seen in the previous chapters in relation to biological cells—especially regarding the description of their behaviours—dynamical systems could represent a general unifying framework also for robotics. This chapter explains the reasons and motivations for which this approach could be promising for robot design.

11.1 Dynamical systems in robot control design

Pfeifer and Scheier in the book “Understanding intelligence” [Pfeifer and Scheier, 1999] present the dynamical systems approach to designing autonomous agents in the “Other approaches” chapter, after the classical methods like the subsumption architecture, neural networks and others.

Dynamical systems (see Chapter 4) is a general mathematical framework, not specific for a particular field. For this reason, there is more than one possibility of applying it to a particular problem domain like that of robotics. Indeed, the metaphors introduced by dynamical systems are suitable for **analysing** and **describing** a robot’s behaviour, but—more interestingly—also for **designing** artificial agents. The generation of behaviour can be, for example, achieved by defining differential or difference equations that govern the robot’s variables; they can also specify the relationships among *sensors* and *internal* variables and eventually these to *actuators* variables. See also Section 12.1.1 for a discussion on the role of attractors (steady states of the system) in robotics.

But, as pointed out by Pfeifer and Scheier, it is much easier to use dynamical systems tools to describe the robot’s observed behaviours rather than write down the equations and integrating them for obtaining the solutions which then guide robotic agents. In this regard, in the following chapters, we will present our contribution which is rooted and based on dynamic systems robotics ¹ and the first successful explorations carried out on the use

¹We will refer to this expression to intend the applications of dynamical systems tools

of Boolean networks (i.e. discrete dynamical systems) as robot controllers.

Chapter 12

On the design of robots with behavioural differentiation capabilities

To overcome the difficulties of robot controller design but at the same time exploit and put into practice the powerful tools and metaphors made available by the dynamic systems framework, we introduced the “attractor landscape” metaphor as a first-class abstraction for robot controller design. In our vision, the attractor landscape concept, especially its topology, aims to be the ultimate goal of any procedure for robot controller design. On the top of it, as we shall see, we could map robot behaviours or let the attractors that compose it describe its behavior through the values of the variables that determine the state vector of the system. Just for example purposes, in the context of Boolean network-based robotic, the attractor landscape can be subject to optimisation or evolution employing any automatic method in order to obtain the desired landscape, appropriate for the given task to be solved.

The additional conceptual step we want to take in the field of robotics is to bring the richness of differentiation dynamics into robotic controllers. The tools and methods developed in the understanding of cell differentiation would allow robots to express various specialised behaviours, the latter triggered by signals or internal dynamics, and the ability to switch between them while maintaining a single controller, just as the cells are capable of it with a unique set of genetic instructions (DNA). The methods developed in the study of differentiation can allow us to model the attractor landscape according to need and at the same time to analyse its resulting complex dynamics to provide us with a composition of behaviours (attractors) suitable for accomplishing complex tasks, hopefully, comparable to those of real biological cells.

In the next sections, we go through the details of our conceptual propos-

als, here only sketched. We also provide some proof of concepts for testing their potentiality and applicability.

12.1 Attractor landscape as metaphor for designing robot controller

From a robotics perspective, GRNs are extremely interesting because they are capable of producing complex behaviours. As we have stressed in many chapters, cell differentiation can be modeled using GRNs and the dynamics of this process can be studied by means of dynamical systems methods. In this scenario, the state of a cell is represented by an attractor in the state space of a dynamical system and the transitions between cell states correspond to transitions between attractors. This view suggests a visionary approach: apply the metaphor of landscape attractor to design specific cell dynamics that can match the attractor landscape required for attaining a target behavior in a robotic system. The constraints prescribed by the robotic application are just the correspondence between behavioural attractors in the robot and cell attractors in the cell, along with specific transitions between attractors. This perspective may lead to applications in bio-robotics and it may also help synthetic biology systems design, which may benefit from methods developed for complex dynamical systems. We believe that this level of abstraction can provide a common vocabulary and a shared set of categories between researchers in robotics and synthetic biology.

With this contribution we proposed some guidelines for making this approach viable, illustrating these concepts with examples and case studies in bio-robotics, case studies will be described in Section 14.1.

The complex behavior exhibited by cell dynamics can be interpreted from a robotics viewpoint, suggesting the possibility of achieving robust and adaptive behaviours in robots—and group of robots—by exploiting the dynamical properties of GRN models. These models can be effectively used as *robot programs*¹. The key motivation of this idea lies in the possibility of applying dynamical system theory to robotics [Pfeifer and Scheier, 1999, Pfeifer and Bongard, 2006, Beer, 1995], exploiting the tight link between artificial intelligence and dynamical systems, that consists primarily in the fact that information processing can be seen as the evolution in time of a dynamical system [Serra and Zanarini, 1990, Bar-Yam, 1997]. The archetypal case of this approach consists in associating the initial conditions of the dynamical system to the input of the problem and let evolve the system in time until it reaches a steady state, which is then interpreted as the output, i.e. the answer to the problem. An example in theoretical computer science is the

¹According to [Russell and Norvig, 2009], we call *robot program* the computational model of the system that maps the percepts of the robot to the actions it takes, possibly according to an utility function and a goal.

solution of the satisfiability problem through Boolean networks [Milano and Roli, 2000], while a typical example in robotics is represented by the different gait patterns in a quadruped robot, each corresponding to one specific attractor in the sensory-motor system of the robot (see [Pfeifer and Bongard, 2006], chap. 4).

Preliminary results in this direction have been achieved in controlling robots by means of Boolean networks. The effectiveness of this approach was demonstrated through experiments on both simulated and real robots [Roli et al., 2011a, Garattoni et al., 2013, Roli et al., 2013, Roli et al., 2015]. These experiments showed that BNs can be successfully used to control robots and therefore that a non-trivial behavior can be attained by a system sharing some similarities with biological cells. The imagination would then run to the synthesis of specific cells controlling micro-robots, produced by synthetic biology (SB) approaches: given the GRN designed *in silico* by means of an automatic procedure,² a synthetic cell is produced by composing elementary cellular bricks. The most natural way to achieve this goal would be either to reproduce a given GRN by means of biological material, i.e. composing a circuit composed of wet logical gates, or to synthesise a cell characterised by a given low-level dynamics, corresponding to the target GRN. Unfortunately, this low-level approach might introduce too many constraints on the design process and turn out to be extremely complicated, if not impossible.³ We believe that a different strategy can be successfully applied, which lies on raising the abstraction level of the analogy from the details of the dynamics to that of *attractor landscape*. Indeed, an in-depth analysis of the GRN-controlled robot dynamics showed that robot’s behavior can be decomposed into elementary behaviours, represented by attractors in the network state space, connected by trajectories that can be controlled by specific inputs.

This result suggests the visionary approach we propose, for the first time, in this contribution: apply the metaphor of landscape attractor to design specific cell dynamics that can match the attractor landscape required for attaining a target behavior for a robot. Indeed, the constraint prescribed by the robotic application is just the correspondence between behavioural attractors in the robot and cell attractors in the cell, along with some specific transitions between attractors. Let us suppose we have to design a micro-robot controlled by a (synthetic) cell—or a populations of cells—whose dynamics in terms of attractors and transitions among them is sufficiently known. A correspondence between cell attractors and robot elementary behaviours can be defined and the chemical signals that force the transitions between cell states can be used as inputs to the robot for changing its ele-

²Preliminary results on automatic design of GRNs for cell differentiation have been recently achieved [Braccini et al., 2017]. Results on automatic design of Boolean networks have also been presented in [Benedettini et al., 2013].

³Anyway, should this be possible in the future, the approach proposed in this contribution would still be useful.

mentary behavior.

12.1.1 Attractors in robotics behavior

The concept of attractor in robotics has been introduced in the context of a dynamical systems approach to design robotic systems in the early '90s (see [Pfeifer and Scheier, 1999], Chap. 9). Here the rationale is the one already proposed in cybernetics and cognitive complex systems, which states that the steady states of the system (i.e. its attractors) represent its typical behaviours. A prototypical example is that of different kinds of gait shown by a robot: despite the fact that the controller is always the same, different environmental conditions influence the final attractor of the system, which converge to a steady state that fits the environment (e.g. walking or trotting).⁴

Along this line are the experiments in *Boolean network robotics* [Roli et al., 2011a, Garattoni et al., 2013, Roli et al., 2013, Roli et al., 2015].

In BN-robotics, the robot is controlled by means of a Boolean network: the value of some nodes of the BN are imposed from the robot sensor readings and the actuators of the robot take the value of some BN nodes. The BN is trained by means of a learning algorithm that manipulates the Boolean functions (and possibly also node connections). The algorithm employs as learning feedback a measure of the performance of the BN-controlled robot (in the following, BN-robot) on the task to perform, such as in evolutionary robotics [Nolfi and Floreano, 2000]. For example, it was shown that a BN-robot can learn a composite mission, in which the first task is to perform phototaxis; then, after a sharp sound is perceived, the robot performs anti-phototaxis [Roli et al., 2011a, Garattoni et al., 2013, Roli et al., 2013, Roli et al., 2015]. A dynamical systems' analysis shows that the behavior of the robot is mainly composed of three attractors: in the first the robot steadily rotate and in the second the robot goes straight. When the frontal light sensor switches on, the BN trajectory exits from a 'rotate' attractor and jumps into a 'go straight' attractor. Subsequently, when the sound is perceived, the trajectory exits from this attractor and moves to a third attractor, the one corresponding to the action 'escape from light'. This dynamics emerges from the learning (evolutionary) process that shaped the BN. The results achieved in BN-robotics are still preliminary, yet quite promising as they show that a GRN model can be effectively used to control a robot that has to attain a non-trivial goal. Further results on GRN models used in robotics are summarised in a survey by one of the authors of this work [Braccini, 2017]. Related to BN-robotics are works in evolutionary robotics, where robots are controlled by artificial neural networks, which are designed by means of evolutionary computation techniques [Nolfi and

⁴See [Pfeifer and Bongard, 2006], page 98.

Floreano, 2000]. An important research line in evolutionary robotics that is quite relevant for BN-robotics and the perspective presented in this work is the one that emphasises the role of *embodiment* in evolved robot [Beer, 1995, Beer and Williams, 2015]. Indeed, the behavior of a BN-robot emerges from the interaction among its sensors and actuators (and the body of the robot itself), the BN dynamics and the environment. In a sense, the experiments in BN-robotics are an instance of the evolution of minimally cognitive behaviours [Beer, 1996, Ziemke, 2005]. The problem of programming and re-programming evolved GRNs has been recently addressed from the perspective of algorithmic complexity and causality [Zenil et al., 2018]. This study proposes a *causal interventional calculus* that makes it possible to steer complex evolved systems. Such an approach may be extremely useful in the context of GRNs controlled robots. For the sake of completeness, we also mention the fact that the automatic design of control software for robots is currently a prominent topic in robotics research, especially when swarms of robots are involved [Francesca and Birattari, 2016].

Following these recent advancements and mainly the achievements in BN-robotics, in the Section 14.1 we will illustrate the use of attractor landscape to bridge robotics and SB.

Chapter 13

Biological Models in Robotics

In order to evolve robots, or sets of robots, capable of increasingly complex tasks, we need to apply to the robotics field more and more powerful models, techniques and methodologies. Natural systems exhibit properties like robustness, adaptiveness, flexibility, scalability and reliability; and they represent a source of interest for the construction of artificial systems. In particular, natural systems like insect colonies and flocking birds exhibit intelligent emergent collective behaviours. We are interested in the dynamical mechanisms at the basis of these systems that lead to the creation of such global level structures, like self-organisation behaviours, from interactions among lower-level components. Complex system science deals with the study of how these low-level parts of a system give rise to the collective behaviours and how the system interact with its environment.

As we have seen in the Chapter 4.3, real living cells show properties typical of critical systems: robustness, adaptiveness and high computational capabilities. Since the optimal balance between robustness and adaptiveness and high computational capabilities are mandatory requirements for attaining complex behaviours also in artificial context, we believe that concepts and mechanisms underlying of the living cells are exploitable, hopefully, to design agent's behaviours as complex as these. As we have already seen, genetic regulatory networks model the interaction and dynamics among genes and they are considered complex dynamical systems able to produce a wide diversity of living cells and organisms. In particular, GRNs describe the complex interactions that ultimately affect the determination of the cellular types. Because of their ability to give rise to different “emerging” cell types depending on their internal dynamics and the external stimuli received, these GRN can be engineered and used to control or to evolve robots.

Here, the most relevant examples of adoption of GRN models in robotic will be presented. In certain case the dynamics of the GRN-based models are used to directly control robots, in others the GRN mechanisms are adopted, similarly to the biological morphogenesis, to develop the robot's neural net-

work control, the robot's morphology or the pattern formation for swarm of robots.

Below, the examples are grouped so as to reflect how the models are used in robotics: automatically evolve a robot controller, automatically design a robot morphology, generate pattern for swarm robotics and automatically co-evolve morphology and controller for robot.

13.1 GRN-based Models for Designing Robot Control

First Example

Eggenberger in the paper “Cell Interactions as a Control Tool of Developmental Processes” [Eggenberger, 1996] suggests that biological concepts as developmental processes are useful and applicable to the field of evolutionary robotics. With the proposed model the length of the genome can be reduced, because no explicit data about the connectivity pattern of the neural net are stored in the genome. The connectivity is not directly encoded in the genome itself but it's mainly determined by the developmental processes. The artificial evolutionary system (AES) includes the following biological concepts and mechanisms:

- Regulatory Units and Transcription Factors, Cell Adhesion Molecules (CAM) and Cell Receptors;
- Cell Differentiation;
- Cell Division;
- Cell Adhesion.

The artificial genome is implemented as a string of integers and it is composed by *regulatory units* and *structural genes*. Regulatory Units are used to activate or inhibit the activity of the structural genes, the latter (if active) modulate the developmental processes producing a substance among these four: transcription factors, cell adhesion molecules, receptors or artificial functions (class that is used to define whether a cell should divide or not).

Regulation of Gene Activity If a cell contains a transcription factor, its code is compared with the code of all regulatory unit in this cell. Depending on a defined *affinity function* the regulatory unit is the activated or inhibited. If a regulatory unit is activated also its structural genes will be activated.

Cell Differentiation Two cells are different if they contain different subset of active genes in the genome. The implemented mechanisms to obtain different cells are: *cell lineage* and *cell induction*. The cell lineage is an autonomous mechanism in which cell differentiation depends on intracellular factors, which are unevenly distributed in different cells. In the cell induction the cells become different because they get different signaling from other cells. To simulate this mechanism the author implemented three different pathways to exchange information between cells: first, there are substances which don't leave the cell and which regulate the activity of genes; second, there are substances which can penetrate the cell wall and activate all cells which are near by; third, there are specific *receptors* on the cell surface which can be stimulated by substances. If a transcription factor has a high enough affinity to the receptor, a gene or group of genes is turned on or off. Only those cells which have a specific receptor on the cell surface will respond to a certain substance. After the process of cell differentiation is finished, the different active genes will determine which substances are produced in a cell.

Cell Division The proposed model is able to simulate cell growth. If the structural genes for cell division in a cell is active, the cell divides itself. The gene activity is dependent, in addition to the affinity function, also on the concentration of the transcription factor. At a certain moment, due to its increased concentration, the transcription factor will turn off the gene for cell division and the growth will stop.

This model has been used to evolve a neural control structure for an autonomous agent. The artificial neurons are the standard ones, with a sigmoidal activation function. As cells can become different, they will express also different substances. To connect two cells or neurons, there are two different types of adhesion molecules and these are stored in lists in the cell. The members of the first list of one neuron are compared with the list of another neuron: if two adhesion molecules of the two different lists have a high affinity to each other a link from the first cell to the second cell is established (if two or more links to the same cell are possible, the substance with the greatest affinity is chosen). The developed neural network has to be linked to the sensors and motors of a real robot, and in order to leave this task to the algorithm implemented, Eggenberger has defined sensory and motor cells with a list of adhesion molecules, in this way other cells can connect to them.

In this work two experiments are presented and the neural controller is evolved by means of a genetic algorithm that is at the basis of the AES. In the first experiment the robot has to accomplish an object avoidance task (the corresponding fitness is increased if the robot sees an object but avoids a collision) and the second task is a phototaxis plus object avoidance (in this case the fitness is the same of the first task, in addition is increased if the

robot moves away from its initial position and if the robot is near the light source). For both tasks the number of initial cells are the same (thus the length of the genome is fixed) and by means of the mechanisms introduced in the AES the neurons can grow and multiply. Therefore, the neural network is evolved and the number of cells is multiplied even though the genome length was fixed.

The purpose of this work is to show that the introduced AES, with the gene regulatory mechanisms, can control the main developmental processes and can evolve functioning neural networks for autonomous agents with number of neurons and patterns of connections not explicitly stored in the artificial genome.

Second Example

Another example of design of neural network controller, in addition to that previously presented, can be found in the following papers: “Morphogenesis of neural networks” [Michel and Biondi, 1995b] and “From the Chromosome to the Neural Network” [Michel and Biondi, 1995a], both of the same authors. In these papers it is proposed a model, inspired from biology, of morphogenesis process with the aim to synthesise an artificial neural network to lead an autonomous robot. Both structure and weights of the neural network are defined by the morphogenesis process. The model was inspired to the biological principle of the *proteins synthesis regulation*¹. The morphogenesis process starts on a single cell enclosing a chromosome and a message list; this list corresponds to the set of proteins available in the biological cell. Cells can divide and establish connections among them by means of a sort of production system that uses and produces messages (representing proteins), through rules. Each rule can be divided in three parts, two conditions and one action part: a set of messages whose absence in the cell message list is necessary for the rule to be fired (corresponding to the biological *repressors*), a set of messages whose presence is necessary for the rule to be fired (corresponding to the biological *activators*) and a set of produced messages which are added to the cell message list when the rule is fired (corresponding to the *biological synthesised proteins*). This morphogenesis process is general: it is able to create any kind of neural network. It allows recurrent connections, different kinds of neurons with different transfer functions and different kinds of links with different learning rules. Thus the space of neural networks explored is theoretically unlimited [Michel and Biondi, 1995a].

To evolve chromosomes classical genetic algorithms have been used with a single point crossover, random mutations and a genetic operator which

¹Protein synthesis is the final stage of gene expression. Once synthesised, most proteins can be regulated in response to extracellular signals and in addition, the levels of proteins within cells can be controlled by differential rates of protein degradation. <http://www.ncbi.nlm.nih.gov/books/NBK9914/>

can add or remove a random element in one of three parts of a rule, or add or remove whole rule at the end of the chromosome. Natural selection uses a binary fitness function (life or death); the evolution is progressive and in order to obtain more complex and powerful systems it is necessary evolve the environment (once a large enough population is able to survive, the environment get a bit more hostile, so that only the elite of this population can survive, and so on, until the individuals develop elaborate behaviours). A simple goal *go towards food* was chosen; a kind of robot *metabolism* was introduced (a variable represents the internal energy of the robot and it is increased each time the robot eats food and decreased when it moves or it is motionless; if it reaches zero the robot “dies” and is eliminated from the population). After around 300 generations of the genetic algorithm, the population of the neural networks evolved were able to produce the desired behaviour, the robots were attracted by food.

This proposed method, inspired mainly to the biological morphogenesis, has demonstrated to be able to produce neural networks (structure and weights) that generate remarkable, even if simple, task (as the attraction by food).

Others Relevant Examples in the Literature

Another interesting example of evolution of a neural network controller using biological principles can be found in the article “Evolving the morphology of a neural network for controlling a foveating retina - and its test on a real robot” [Hotz et al., 2003]. The proposed model combines artificial evolutionary techniques with bio-inspired developmental processes in order to evolve a neural network that acts as an artificial foveating retina (that is, move the “eye” in such a way that an incoming peripheral sensory stimulus falls in the center of the eye, the eye has to learn to foveate on the stimulus). In particular, this system exploits mechanisms like gene regulation and developmental mechanisms like cell division, axonal outgrowth, synaptogenesis and learning for controlling the structure of the neural network and the synaptic weights. After the simulation, the evolved controller was tested in a real robot arm equipped with a CCD camera and the arm has proved to be able to foveate considerably well.

In the paper “Harnessing Morphogenesis” [Jakobi, 2003] is presented another fascinating example of controller design making use of a biologically inspired developmental model. This system is able to develop a multicellular organism, starting from a single cell, exploiting similarities with biological morphogenesis. The behaviour of the cell, during development, is controlled by a GRN that can be thought as a dynamical system. The product of this developmental process is interpreted as a recurrent neural network robot controller; this model was able to evolve controllers for accomplish a corridor following and an object avoidance task.

The paper “Evolving Embodied Genetic Regulatory Network-Driven Control Systems” [Quick et al., 2003] presents experiments in which a GRN-based controller is embodied in artificial organisms. In this model, called *Biosys*, the interplay between the dynamics of the embodied GRN controller (the suitable genome is evolved through a genetic algorithm) and the environment gives rise to coherent observable emergent behaviours. It’s presented a successful experiment of a simulated robot, guided by its GRN controller, able to fulfil phototaxis, but the key point of this work it’s to remark the importance of the role of the environment in the generation of observed behaviour; the environment “selects” the cell dynamics able to produce the desired behaviour.

13.1.1 BN-based robot controller

In the paper “On the Design of Boolean Networks Robots” [Roli et al., 2011b] it was presented the use of Boolean networks for controlling robot’s behaviour. The approach proposed consists in using one or more BNs as robot program so that the robot dynamics can be described in terms of trajectories in a state space. The authors propose a design methodology based on metaheuristics in which the design of a BN is modelled as a constrained combinatorial optimisation problem: the algorithm manipulates the decision variables which encode structure and Boolean functions of a BN. A complete assignment to those variables defines an instance of a BN. This technique uses an evaluator that produces an objective function value that represents the performance of the current BN and the feedback to the metaheuristics algorithm, that, in turn, proceeds with the search. Another possible way, suggested in this article and that can be combined with the previous presented, to design the BN for a robot program is to exploit its dynamics in order to satisfy given requirements. For example, the attractors with largest basins of attraction may correspond to the high-level robot’s behaviours and the transitions between attractors to the transitions between behaviours.

The case study presented consists of a robot that must be able to perform two different behaviours: going towards the light (phototaxis) and subsequently moving away from it (antiphphototaxis) after perceiving a sharp sound (like an hand clapping). The environment, in which the robot is simulated and later tested in reality, consists of a square ($1m \times 1m$) with a light source positioned in one corner. The robot, in the beginning of the experiment, is located in random position close to the opposite corner of the arena with respect to the light and the performance measure used to evaluate the robot behaviour is an error function that has to be minimised (smaller is the error, better is the robot performance). The BN implementing the robot program is subject to a synchronous and deterministic update and the number of network nodes is has been set to 20 (sensors and actuators have been mapped onto some node of the BN). The Boolean network was designed with a lo-

cal search techniques, that is a simple stochastic descent in which a move can change one value in a node function's truth table; a random entry in the truth table of a randomly chosen node is chosen and accepted if the corresponding BN has an evaluation not worse than the current one. The initial connections among nodes are randomly generated with $K = 3$ (no self-connections) and are kept fixed during the search; the initial Boolean functions are generated by setting the 0/1 values in the truth tables uniformly at random. The BN-robot is trained in two sequential phases: in the first, the learning feedback is an evaluation of the robot's performance in achieving only phototaxis and in the second the performance measure takes into account both the phototaxis and antiphototaxis. In this way, it becomes possible to study the properties of the evolution of the BN-robot when its behaviour must be adapted to a new operational requirement.

The results obtained from this experiment were presented in the paper [Roli et al., 2012]. Analysing the dynamics of BN-robots trained, using concepts of dynamical systems theory and complexity science, they have found that the successful performing robots, which show the capability of robustly attaining the learned behaviours while adapting to new tasks to perform, are characterised by both number of fixed points and complexity higher than those of unsuccessful ones. The number of fixed points is an indicator of the generalisation capabilities of the system as they represent micro-behaviour which are combined to achieve a global behaviour and the measure of the complexity used is the LMC complexity². These results are in accordance with the conjecture that artificial systems able to balance robustness and evolvability work at the border between order and chaos as the living systems, an example are cells.

In the papers "A Developmental Model for the Evolution of Complete Autonomous Agents" [Dellaert and Beer, 1996] and "Co-evolving Body and Brain in Autonomous Agents using a Developmental Model" [Dellaert and Beer, 1994] it is presented a model for neural development in which a random Boolean Network is used as an abstraction of the genetic regulatory network inside a cell. The introduced developmental process has showed to be able to successfully evolve agents than can execute simple tasks (i.e. line following).

13.2 GRN-based Models for Evolving Robot Morphology

The paper "Evolving Morphologies of Simulated 3d Organisms Based on Differential Gene Expression" [Eggenberger, 1997] reports a biologically inspired model used to evolve 3d shapes of simulated, multicellular organisms. The model has the same concepts and biological mechanisms introduced in

²LMC complexity is defined as $C' = HD$, where H is the *entropy* and D is the *disequilibrium* of the BN states in the trajectories.

the previously presented article of the same author [Eggenberger, 1996], in addition introduces the *positional information* and pattern formation in development. With this last mechanism the cells acquire positional identities as in a coordinate system and then interpret this information according to their genetic constitution and developmental history. An example of such mechanism is a concentration of gradient of a *morphogen* which every cell is able to read. A morphogen is a substance governing the pattern of tissue development in the process of morphogenesis, and the positions of the various specialised cell types within a tissue³. In the implementation proposed by Eggenberger the morphogen is just a kind of transcription factor (TF) which can diffuse to other cells and can change the state of some genes in cells able to read this message. This mechanism, already implemented by the regulatory mechanism in the AES, is not just a simple signalling, because the reading mechanism (the *cis-regulators*, which are binding sites for transcription factors) is controlled by the AES. Therefore the same morphogen can have very different effects on different cell. Some examples of such effects are changes in cell type, cell division rate or motility. Using the proposed

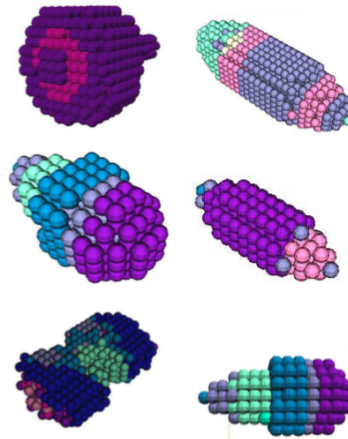


Figure 13.1: Examples of evolved forms by means of the AES; the fitness function evaluated only the number of the cells and the bilaterally of the found organisms. Image taken from [Eggenberger, 1997].

AES, Eggenberger was able to evolve three dimensional shapes that could be used as projects for three-dimensional robot. Some examples, taken by the original paper [Eggenberger, 1997], of these evolved forms are represented in Figure 13.1.

³<https://en.wikipedia.org/wiki/Morphogen> Date: 04/01/2016

13.3 GRN-based Models for Pattern Generation in Swarm Robotics

Inspired by the biological morphogenesis and the evolution and structure of **networks motifs**, in the paper “Evolving Network Motifs based Morphogenetic Approach to Self-Organizing Robotic Swarms” [Meng and Guo, 2012] it’s presented a GRN-based control model. This model has the aim to autonomously generate dynamic pattern for swarm robot in complex environment. Network motifs are pattern of interconnections occurring in complex networks at numbers that are significantly higher than those in randomised networks [Milo et al., 2002]; therefore they represent building blocks for most complex networks. The authors propose a developmental method where the artificial GRN-based controller will be automatically evolved by an evolutionary algorithm using some predefined network motifs as basic building blocks.

The aim of this model is to generate suitable shapes so that swarm of robots (with limited sensing and communication capabilities) can traverse an unknown environment with various constraints. Inspired by the biological morphogenesis, in which the morphogen gradients are either obtained from the mother cells or generated by a few cells known as organizers, an organising robot is selected in order to generate the final target shape (considering the current environmental constraints) for the swarm robot. The regulation of gene expression is used to model the base concepts of the general GRN-framework. This framework will be embedded into each robot of the system, but only the organising robots will activate the framework and generate the suitable shapes virtually in its own mind. Then, the generated shape will be sent to all the other robots through local communication so that they can merge to this shape automatically. In the framework, the transcription factors (TFs) are used to denote the input of the GRN framework: TF1 measures the minimal distance from the current robot to the nearest obstacle and TF2 is used to maintain the number of robots. Two genes, G1 and G2, can be thought as the processors of the robots: they are responsible to process the inputs of the GRN-framework and send signals to trigger the outputs. Three proteins represent the output (actions) of the framework: P1 grow into an area; P2 skip an area and try to grow into another area; and P3 stop growing. Five basic network motifs, that represent the regulations (building blocks) to constructing the GRN-framework, are proposed: *positive*, *negative*, *OR*, *AND* and *XOR* ⁴. Then, using the predefined network motifs as building blocks, an evolutionary algorithm is applied to evolve structure and parameters of the GRN-framework. In this manner each link in the GRN framework can be modeled by one of the basic net-

⁴The precise mathematical formulation of these types of regulations can be found in the paper [Meng and Guo, 2012].

work motifs. By means of the evolutionary algorithm we need to optimize the parameters of the general GRN-framework in order to instantiate a GRN framework able to generate suitable shapes for swarm robots to adapt to unknown environments (therefore the fitness function depends on the distance from nearest obstacle and from the number of the robots within the shape). The evolved GRN framework, presented in this work, is able to generate the final target shape, starting from a single robot, for different environments. Moreover, a case study is conducted in which the robots have to traverse a complex unknown environment with different constraints along the path. The swarm of robots was able to adapt its shapes, using the GRN-based framework developed, during the traverse of a complex environment.

In the paper “Evolving Hierarchical Gene Regulatory Networks for Morphogenetic Pattern Formation of Swarm Robots” [Oh and Jin, 2014] it is presented an approach to pattern formation for swarm robots, inspired by biological morphogenesis, that uses a hierarchical gene regulatory network (EH-GRN) evolved using network motifs.

An interesting European project relating to swarm of GRN-controlled agents whose goal is collectively organise themselves into complex spatial arrangements is Swarm-Organ (visit <http://www.swarm-organ.eu/>).

13.4 GRN-based Models for Co-Evolving Body and Brain

In “Evolving Complete Agents using Artificial Ontogeny” [Bongard and Pfeifer, 2003] it is presented an artificial evolutionary system, *Artificial Ontogeny* (AO), that combines an ontogenetic development with a genetic algorithm in order to evolve complete agents, that is both the morphologies and controllers of robots. Each genome, evolved using a genetic algorithm, is treated as genetic regulatory networks, in which genes produce gene products that either have a direct phenotypic effect or regulate the expression of other genes. In this model there is a translation from a genome (genotype) into a three-dimensional agent (phenotype), later evaluated in a physically-realist virtual environment, that takes place via ontogenetic processes: the differential gene expression and the diffusion of gene products transforms a single structural unit into an articulated three-dimensional multi-unit agent composed of several structural units that can contain sensors, actuators and a neural network structure. Therefore, each agent begins its ontogenetic development as a single structural unit; structural units (spheres) which are the basic building blocks from which the agent’s morphology is constructed. Depending on the concentrations of gene products within a unit, the unit may grow in size and even split into two units. Each structural unit contains at most six joints (to which other units can attach to them), a copy of the genome and six diffusion sites. Each diffusion site contains zero or more dif-

fusing gene products and zero or more sensor, motor and internal neurons. Three types of sensor can be embedded, by the artificial evolution, within a structural unit: touch sensors, proprioceptive sensors and light sensors. The neurons at a diffusion site may be connected to other neurons within the same unit or in other units. After a unit splits from its parent unit, the two units are attached with a rigid connector. In addition to the morphology of the agent, neural structure may grow within the developing agent. Each genome of population is represented by 100 floating-point values (between 0.00 and 1.00) and is scanned by a parser in order to find the promoters; promotor sites indicate the starting position of a gene along the genome. During the growth phase, the genes may emit gene products: the gene products are treated as chemicals which spread to neighbouring diffusion sites, and to a lesser degree, into neighbouring structural units. There are 24 different types of gene products: 2 affect the growth of the unit in which they diffuse, 17 affect the growth of the agent's neural network and 5 have no phenotypic effect, but rather may only affect the expression of other genes (enhance or repress). In the AO a cellular encoding has been incorporated to achieve the correlated growth of morphology and neural structure. Cellular encoding is a method for evolving both the architecture and synaptic weights of a neural network, by starting with a simple neural network (embedded in each new structural unit) and iteratively applying a set of graph rewrite rules to transform it into a more complex network. If the concentration of one of the 17 gene products, responsible for neural development, at a diffusion site exceeds a concentration of 0.8, and there is a neural structure at that site, the corresponding rewrite rule is applied to the neural structure. This neural development scheme is able to evolve dynamic, recurrent neural network that propagate neural signals from sensor neurons to motor neurons. In order to evolve the complete agents a genetic algorithm (with mutation and crossover) is applied with 200 generations and a population size of 200. Each genome is evaluated (according to a task-specific fitness function) as follows: the genome is copied into a single structural unit and placed in a virtual, three-dimensional environment; morphological and neural development is allowed to proceed for 300 time steps; after this the neural network is activated and the agent is allowed to operate in its noisy environment for 1000 time steps. The agent is then regrown and re-evaluated nine more times, and the agent's fitness values are averaged. Using the AO system, agents able to perform directed locomotion and block pushing (see Figure 13.2 for an example) in a noisy environment were evolved. Another example of a GRN-based evolution of complete autonomous agents can be found in [Dellaert and Beer, 1996]; but in this last model the nervous system develops after the development of the agent morphology.

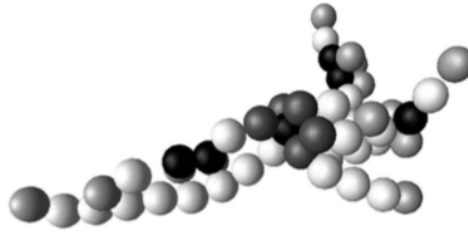


Figure 13.2: An example of evolved agent for block pushing. Image taken from [Bongard and Pfeifer, 2003].

13.5 Morphogenetic Robotics

The term **morphogenetic robotics** has been first introduced in the paper “Morphogenetic Robotics: An Emerging New Field in Developmental Robotics” [Jin and Meng, 2011]. Morphogenetic robotics is an emerging new field in developmental robotics that consist of a class of methodologies in robotics for designing self-organising, self-reconfigurable and self-repairable single or multi robot systems, *using genetic and cellular mechanisms governing biological morphogenesis* [Jin and Meng, 2011]. Biological morphogenesis is the biological process in which cells divide, grow and differentiate, and finally resulting in the mature morphology of a biological organism. Morphogenesis is under the governance of a developmental gene regulatory network and the influence of the environment [Gilbert and Barresi, 2016]. They categorize these methodologies into three areas:

- *morphogenetic swarm robotic systems*: deal with the self-organisation of swarm robots using genetic and cellular mechanisms underlying the biological early morphogenesis;
- *morphogenetic modular robots*: modular robots adapt their configurations autonomously based on the current environmental conditions using morphogenetic principles;
- *morphogenetic body and brain design for robots*: include the developmental approaches to the design of the body or body parts, including sensors and actuators and/or design of the neural network-based controller of robots. The neural structure is the product of neural morphogenesis (*neurogenesis*).

The authors claim that the *developmental robotics* should include both *morphogenetic robotics* and *epigenetic robotics* ⁵: the first is mainly concerned with the physical development of the body and neural control, whereas the

⁵For a comprehensive survey of Epigenetic Robotics, and more in general, of Developmental Robotics see [Lungarella et al., 2003].

second focuses on the cognitive and mental development. The body morphology, as well as the neural structure of the robots is a result of morphogenetic development, on which mental development is based through interaction with the environment [Jin and Meng, 2011]. In Figure 13.3 we can see the relationship between morphogenetic robotics, epigenetic robotics and developmental robotics. The authors of this paper introduce these three categories; below

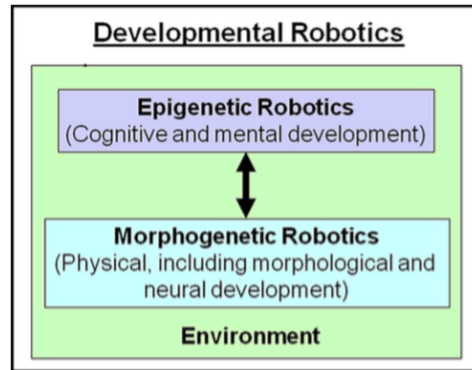


Figure 13.3: Morphogenetic and Epigenetic Robotics are closely coupled not only directly in that the body plan and nervous system are the basis of cognitive development, but also indirectly through the environment. Image taken from [Jin and Meng, 2011].

briefly summarised.

Morphogenetic swarm robotics A swarm robotic system is a multi-robot system consisting of a large number of homogeneous simple robots. In order to apply genetic and cellular mechanisms in biological morphogenesis to self-organised control of swarm robots, it is necessary establish a metaphor between a cell and a robot. The movement of each robot can be modelled by the regulatory dynamics of a cell. In particular Guo Meng and Jin (in some works [Guo et al., 2009, Meng et al., 2013]) have described the movement dynamics of each robot by means of a GRN model, where the concentration of two proteins represents the position of a robot and the concentration of another protein represents its velocity. In this gene regulatory model, the target shape information is provided in terms of morphogen gradients. This morphogenetic approach to swarm robotic systems has the advantage that the target shape can be embedded in the robot dynamics in the form of morphogen gradients, in this way the GRN model can generate implicit local interactions rules automatically to generate the global behaviour. In addition, this model is robust to perturbations in the system and in the environment.

Others examples of morphogenesis-inspired models for swarm robotics is presented in 13.3.

Morphogenetic modular robots Self-reconfigurable modular robots consist of a number of modules. They are able to adapt their shape by rearranging their modules to changing environments. Each module has its “body” and its controller and each can be seen as a cell. In fact, there are similarities in control, communication and physical interactions between cells in multicellular organisms and modules in modular robots. The control, in both cases, is decentralised and the global behaviour emerges through local interactions of the units. Therefore, it is a natural idea to develop control algorithms for self-reconfigurable modular robots using biological morphogenetic mechanisms [Jin and Meng, 2011]. The authors present an example, taken from [Meng et al., 2010], of morphogenetic approach to designing control algorithms for reconfigurable modular robots. Similar to morphogenetic swarm robotic systems, each unit of the modular robot contains a chromosome consisting of several genes that can produce different proteins; the proteins can diffuse into neighbouring modules. The target configuration of the modular robot is also defined by morphogen gradients. Morphogen gradient that each module is able to modify in order to attract or repel neighbouring modules and so adapt the global configuration to the environment or task. The attraction and repellent behaviour of the modules are regulated by a GRN-based controller. Particularly, it is used a hierarchical approach to self-reconfiguration of modular robots: one layer defines the desired configuration of the modular robots while the other layer organizes the modules autonomously to achieve it. This hierarchical structure is similar to those of the biological gene regulatory networks [Jin and Meng, 2011]. This hierarchical controller, inspired by the embryonic development of multi-cellular organism, is resulted efficient and robust in reconfiguring modular robots to adapt to the changing environment.

Morphogenetic body and brain design for robots This category, according the authors, comprises models for neural ⁶ and morphological development in designing intelligent robots. It is also important reproduce the natural co-evolution of development of body and brain in which the cognitive and mental development is influenced by the morphological development (and by the environment) and vice versa. In the paper [Jin and Meng, 2011] it is cited an example of co-evolution in development of robot hand morphology and controller, see Figure 13.4. In this way the shape, the number of fingers and finger segments can be evolved together with their controller in a task-dependent way: different hand morphologies will emerge by evolving the system for different behaviours [Jin and Meng, 2011].

Another example of morphogenesis-inspired co-evolution of body and

⁶The analogy between neural development and biological neurogenesis is here reported but there is no a truly morphogenesis process because the evolution of the formal model which controls the robot is simulated.

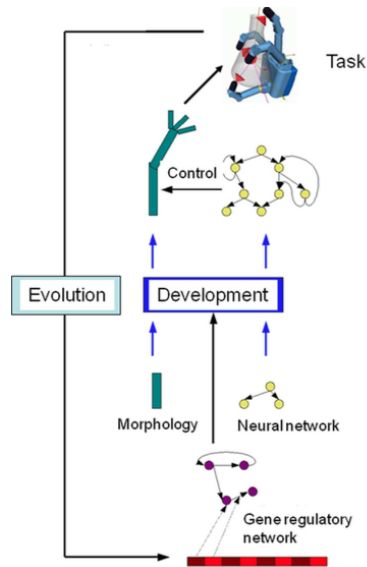


Figure 13.4: Conceptual diagram for coevolving the development of arm and control. Image taken from [Jin and Meng, 2011].

brain is presented in 13.4.

An example of a morphogenesis-inspired body development is presented in 13.2.

13.6 Conclusion

The work synthesised in the previous sections represents a survey of the most relevant examples in the literature concerning the application of genetic regulatory network models in robotics. Examples of GRN-based models are presented for designing robot control, for evolving robot morphology, for pattern generation in swarm robotics and for co-evolving body and brain. Moreover, an introduction to another kind of classification, inspired by the biological morphogenesis, is given with the Morphogenetic Robotics.

Some of the introduced models are biologically plausible, in other words the ideas introduced in the developed artificial system can be equated with the relative biological concepts. Others instead exploit the similarities with the cell's mechanisms but they prefer to be more abstract and therefore more computationally tractable. The majority of the examples found in the literature concerns the design of robot controller, and this is reflected in the paper. This is so far due, in part, to the lack of technological support (modular robotics is an example) and to the needed computational resources (see co-evolution of the body and brain).

Nevertheless, this approach has proven to be an efficient tool for the evo-

lution of robot capable of interesting not trivial behaviours and it represents a field of study with not yet fully explored potential.

Chapter 14

Robotics applications

14.1 Case study I – Attractor landscape: proofs of concept

The abstraction of attractor landscape is the space where robotics meets Synthetic Biology. In this section we illustrate this vision by discussing two paradigmatic examples in which a genetic network is used as control software for robots.

14.1.1 BN model of a simple genetic network

An illustrative example of BN modelling the basic cellular states of a cell is provided by Huang in [Huang and Ingber, 2000]. This BN is a minimalistic example of a biologically plausible GRN, as the genes regulating functions consist of Boolean encoding of relations that can be typically found among genes in real cells. Here we introduce the model and discuss the properties that are relevant for the purpose of this contribution. The network is composed of four genes, named A , B , C and D . In Figure 5.1 (in Chapter 5) the relations among genes and their functions are illustrated.

The state of the network is given by a binary vector of four components, representing the activation state of the genes. For example, state 0001 represents a situation in which genes A , B and C are inactive, while gene D is active. The network is supposed to update its node synchronously, therefore—in absence of external perturbations—each state has a unique successor. Under this updating scheme, the dynamics of the network starting from any initial state (i.e. gene activation profile) is a *trajectory* composed of a transient—if any—and a cyclic *attractor*, which may be a degenerate cycle involving only one state, i.e. a *fixed point*.

The graph representing all the possible transitions between network states is depicted in Figure 14.1. We observe that the dynamics is characterised by four attractors: three fixed points 0000, 0100 and 1110 and a cycle of

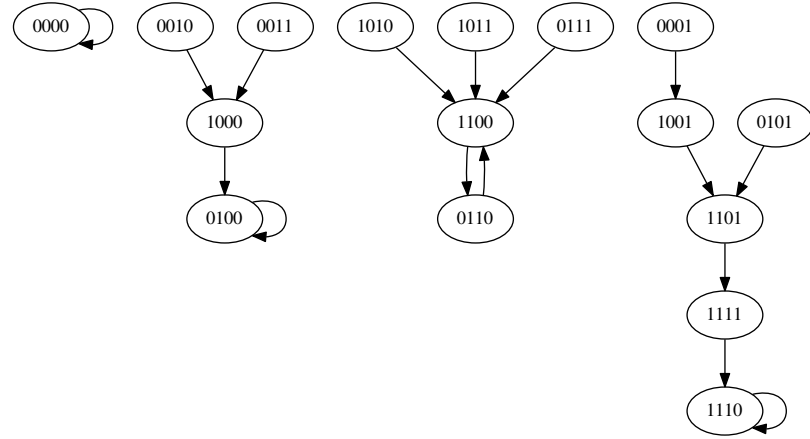


Figure 14.1: State graph of the network defined in Figure 5.1. Note the four attractors: three fixed points 0000, 0100 and 1110 and a cycle of period 2 (1100, 0110).

period 2, (1100, 0110). The attractors represent the main cell states, as they constitute the steady states in the dynamics of the cell. For this reason, they assume a particular importance, as also shown in the original example by Huang, who associate one specific cell behavior to each attractor. The state space (i.e. the space of all possible gene configurations) can be partitioned into *basins of attraction*, each containing all the states that, if assumed as initial condition, leads to one specific attractor. For example, the basin of attraction of the cyclic attractor (1100, 0110) is composed of the states {1010, 1011, 0111, 1100, 0110}.

In absence of perturbations, after a (possibly empty) transient, a cell rests in one attractor. However, when the network in an attractor is perturbed, it might exit from the basin of attraction of the current steady state and move to another one. Usually, in these models a perturbation affects just one node at a time [Serra et al., 2010], therefore it is possible to draw the *attractor graph*, which represents the possible transitions between steady states. The attractor graph of the example we are discussing is depicted in Figure 14.2. The graph is obtained by perturbing each node of each attractor and connecting attractor α to attractor β with an arrow from α to β if the perturbation in α produces a trajectory ending in β —or, equivalently, if the perturbation on α produces a state in the basin of attraction of β . In the case of the cyclic attractor of period two, we numbered the states and denoted by a subscript the perturbed genes as a function of the state. We can observe that it is possible that the same gene, if perturbed, leads the trajectory to

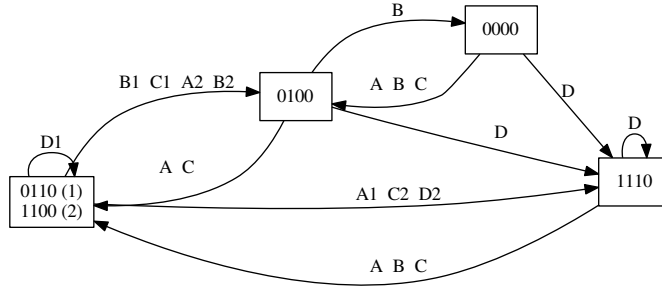


Figure 14.2: Attractor graph of the network defined in Figure 5.1. For clarity, the transients are omitted and only macro transitions between attractors are depicted. A transition occurs after the transitory flip of the value of a gene. The labels on the edges denote the genes which, if flipped, cause the transition.

different steady states depending on the attractor state in which the gene is perturbed. For example, gene *A* leads to attractor 1110 if perturbed in state 0110 and to attractor 0100 if flipped in state 1100.

The network described above models a typical case of cell dynamics and it was used in [Huang and Ingber, 2000] to illustrate the notion of attractors in cell dynamics. In the following subsections we show how this network can be used to control a robot performing minimal yet not trivial cognitive task. The key idea is that attractors are associated to robot behaviours, in the same way as they represent cell behaviours in the biological interpretation.

14.1.2 Example 1: Controlled phototaxis

In this first example, we exploit the properties of the attractor landscape to control the speed of a robot performing phototaxis, i.e. moving towards a light source. Here and in the following case studies we have directly introduced a mapping between attractors and robot behaviours. However, this mapping can be the result of an adaptive process, as indeed done in nature where the interactions between a system (e.g., a cell or even an organism) and the environment emerge as an adaptive process that exploits some regularities in the environment. This process is analogous to the emergence of sensors in nature, where regularities, correlations and sufficiently robust patterns are captured by organisms' parts that assume the role of sensor devices—see [Olsson et al., 2006, Cariani, 1992, Cariani, 1993, Balakrishnan and Honavar, 1996] for a discussion on the evolution of sensors, both in nature and in robotics. Intermediate situations are possible between these two

extreme possibilities, such as in the case studies in BN-robotics that we have previously mentioned (see Section 12.1.1). In those BN-robots, some nodes of the network are directly connected to a sensor (e.g. a light or a proximity sensor) and their value is imposed by sensor readings and actuators are directly controlled by the values of some predefined node. Despite this *a priori* setting, nothing is imposed on the way the network will use the information set on its inputs nor the way it will control the robot actuators, as the connections among nodes and node functions are the result of an evolutionary process. In a sense, we may say that this evolutionary process defines the semantics of the information received and elaborated by the robot.

The attractors of the network are characterised by a different number of active genes, from 0 to 3; this property can be easily exploited as a control factor for the speed of the robot: the more the number of active genes in the state, the higher the speed of the robot. The control genes are D , which is temporarily switched on when the robot sees the light, and gene B which temporarily is deactivated whenever the luminescence gradient perceived by the light sensors exceeds a given threshold. As an aside comment, we observe that, whilst we are using the terminology typical of robotics, we are just describing a dynamical system interacting with the environment, like a cell. The network starts in attractor 0000, which represents the quiescent state where robot's wheels do not move. When the robot perceives the light, gene D is switched on—as if it was activated by an external molecule. At each control step of the robot, the network updates its state; therefore, after the perturbation occurring on gene D , the network enters the basin of attraction of fixed point 1110, which is reached in few steps. Then the robot moves towards the light and progressively slows down, as an effect exerted by gene B , which is temporarily suppressed (i.e. set to 0) as soon as the light intensity detected exceeds a fixed value. Eventually, the robot stops when it is close to the light source. Note that the stop state corresponds to fixed point 0000, which is reached from attractor 0100 just by setting B to 0. The video of a representative run is available online at <http://www.lia.disi.unibo.it/~aro/download/attractors/> as `video-01`. The same network can be used to control a group of robots performing the same task. We performed this and the following experiments in a simulated environment by the means of ARGoS [Pinciroli et al., 2012], which is one of the most widespread robotics simulators. The main steps of this dynamics are depicted in Figure 14.3 and a video of the simulation is available as `video-02`.

In case this network is used to control the behaviour of a swarm of robots, one may want to attain a final situation in which robots are evenly distributed across the light sources, similarly to clustering phenomena in cell biology. To attain this goal, the very same network can be used and gene D is activated as long as the robot density perceived by a robot (through its proximity sensors) exceeds a given threshold. In this way, the temporary activation of gene D moves the network to the attractor corresponding to the maximal

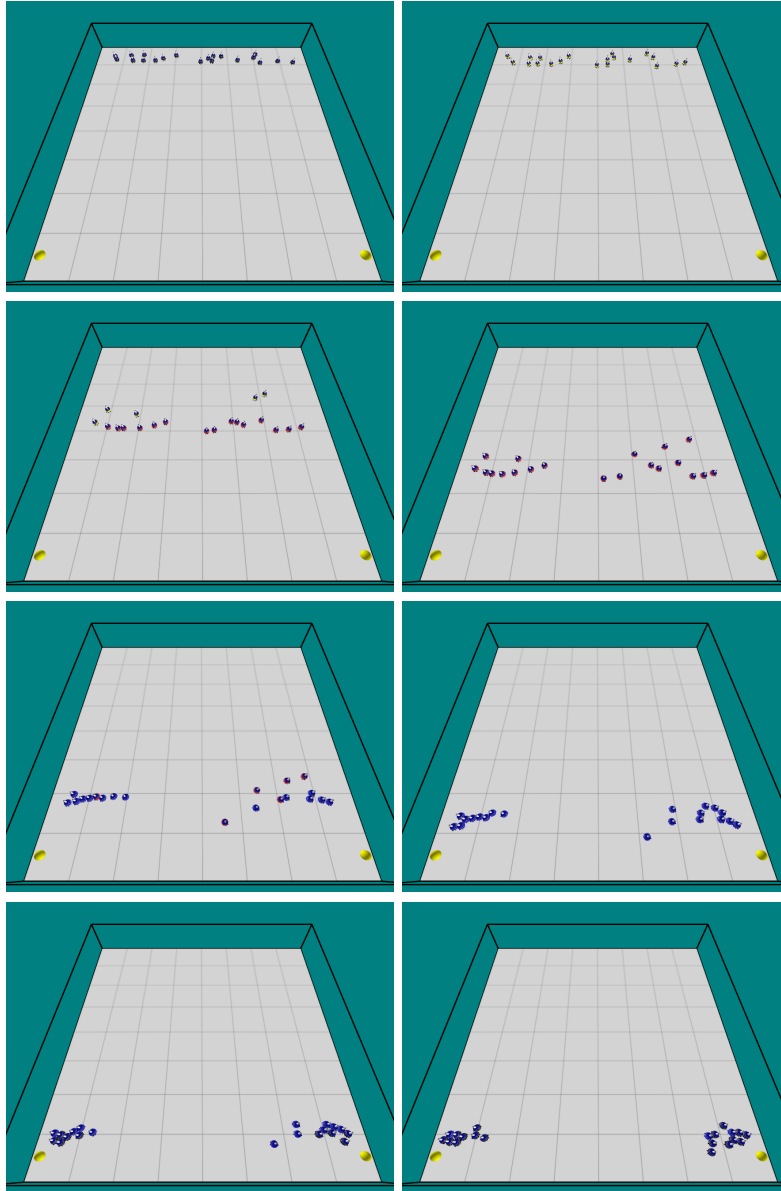


Figure 14.3: Main phases of the phototaxis behaviour of a group of robots (from top to bottom and left to right). Robot colours denote their attractor (and consequently, their speed): black \rightarrow 0000, yellow \rightarrow 1110, red \rightarrow (0110, 1100), blue \rightarrow 0100.

speed so that the robot has the chance to move and find another less crowded light source. The main phases of this dynamics are depicted in Figure 14.4, while the video of a typical simulation is available as `video-03`.

Although this has been introduced purely by way of example, we reserve the right to make a comparison of the proposed approach against other techniques [Amé et al., 2006], present in the literature, concerning robot swarm aggregation in the future.

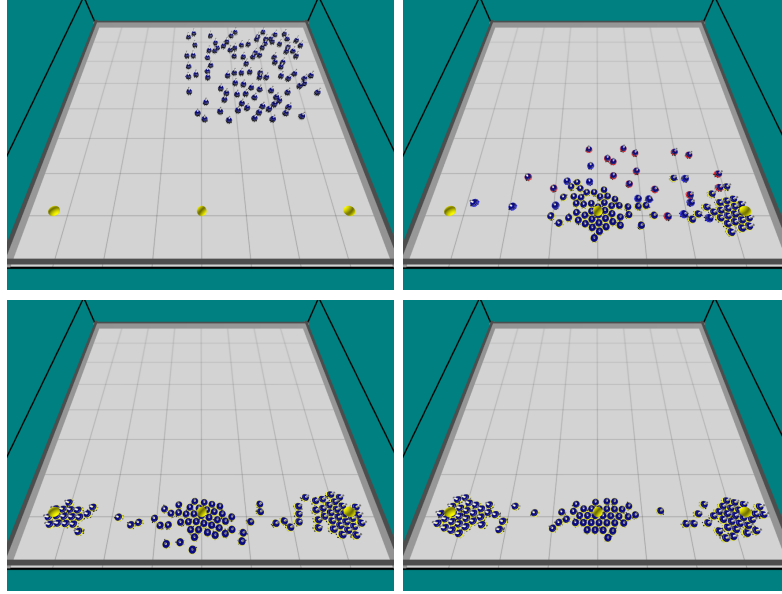


Figure 14.4: Main phases of the phototaxis behaviour of a group of robots (from top to bottom and left to right), trying to gather around a light source so as to split into approximately equal groups. Also in this case, robot colours denote their attractor (and consequently, their speed): black \rightarrow 0000, yellow \rightarrow 1110, red \rightarrow (0110, 1100), blue \rightarrow 0100. Note that robots in a dense group are coloured yellow, i.e. they are moving at a high speed. Therefore, in this case the equilibrium reached at the end of the run is dynamics, rather than static.

14.1.3 Example 2: Actions triggered by an external stimulus

As a second example of the use of the dynamical properties of a cell model, we show an alternative approach to encode inputs and outputs in the network. In the previous example, the mapping between cell model and robot has been achieved by temporarily setting some gene to a specific activation state and using the entire genetic profile to decide the actions the robot should take (in the previous case, the speed of the robot). Another approach consists in directly connecting some genes of the network to external inputs—i.e. treat

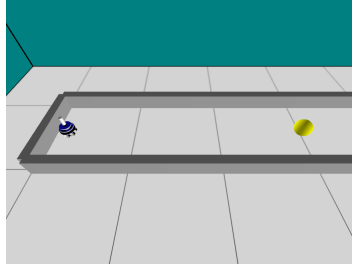


Figure 14.5: Corridor scenario: the robot points towards a goal and its movement triggered by an external stimulus.

them as receptors—and using the value of some specific genes to directly control some low-level robot actions. In the example we discuss here, we consider a simple scenario in which a robot is placed in a corridor that has to be traversed so as to reach a target. In Figure 14.5 the initial situation is depicted; the target is represented by a light source, but it may be any source of a signal that the robot can perceive, such as sound or temperature. On the biological side, this source can be any chemical source and the phenomenon would be chemotaxis. The network controlling the robot is the same as the one used in the previous example, just with a different encoding of inputs and outputs. Here we suppose that an input gene is clamped to 0 or 1 as a consequence of an external signal; when a network node is forced to a constant value, the network state graph changes and some transitions (along with some states) do not longer exist. To adhere the biological framework depicted by Huang, we suppose that signal exert their effect on the network to condition the transitions from an attractor to another one; in this way, attractors still represent the main behaviours of the robot and the transitions between them are achieved by clamping a node to a constant value so as to control the transient from an attractor to another one. In the scenario we discuss in this example, the control gene is again D and the output gene is A , which acts as a binary selector: if $A = 0$, then the robot holds, otherwise it moves straightforward. The initial state is the quiescent one (0000) and, when an external signal is performed (e.g. a sound) and during the time interval it is perceived by the robot, gene D is clamped to 1. As shown in Figure 14.6, as soon as D is set to 1, the network state moves to 0001 and, while $D = 1$, the network trajectory eventually reaches 1111, which is a fixed point as long as $D = 1$. We may call this particular steady state a *conditional attractor*, i.e. an attractor conditioned to an external conditioning on some genes, to distinguish this case from that of *original attractors* which are the ones characterising the *autonomous*¹ dynamics of the network. Once this conditional attractor is reached, the external stimulus can be detached

¹In the context of dynamical systems, an autonomous system has no inputs and it is subject to an internal dynamics.

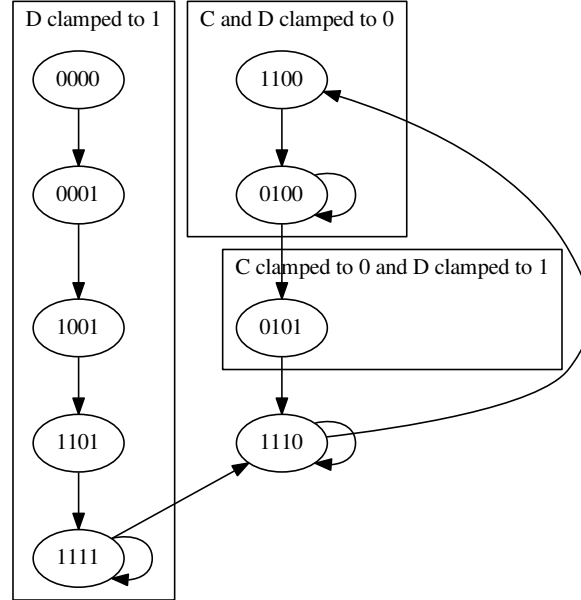


Figure 14.6: State graph of the network controlling the robot. Whilst the network is the same as the one in the previous example, its state graph is different because some genes are clamped either to 0 or 1 from an attractor until a new attractor is found.

from D and the network freely reaches the original fixed point 1110. Along this trajectory, gene A is always 1 and so the robot moves straight. The possibilities opened by clamping one or more genes to a specific value until a new attractor is reached make it possible to introduce also a stopping condition to this behavior: when we want the robot to stop, both C and D have to be clamped to 0 and so after two steps a new conditional attractor is reached with $A = 0$ and the robot stops. At this point, the plasticity of the network enables us to control again the movements of the robot toward the light source by keeping C clamped to 0 and activating or inhibiting D , which then will act as a switch to make the robot moving and resting.²

²Videos of these behaviours can be watched at <http://www.lia.disi.unibo.it/~aro/download/attractors/> as video-04 and video-05.

14.1.4 Conclusion: Implications on robotics and synthetic biology

We believe that the notion of attractor landscape provides an effective abstraction level for cross-fertilisation between robotics and SB. On the one hand, robotics may exploit advances in SB so as to devise unconventional control systems. Indeed, the examples we have presented in the previous section illustrate a viable approach to combine robotics and SB, which consists in exploiting synthetic cellular circuits to control robots. This “understanding by building” cross-discipline methodology can produce unforeseen developments in both fields; indeed, results obtained from the evaluation—in simulation or in real world—of robots designed exploiting these cellular synthetic bricks may provide biological insights and hypotheses to motivate new experiments, that in turn may lead to the construction of new bricks. In addition, this approach opens the possibility of designing and building hybrid robots, made also of biological components. Typical scenarios of such creatures are environments where human exploration is not possible, such as oceans and human and animal body and also plants, where swarms of micro-robots may collectively accomplish a mission. On the other hand, the design of synthetic cellular systems may be formalised in terms of an embedded agent perceiving the environment and acting on it—as done in robotics—and design techniques for control software in robots may be used in SB design.

We are aware that the approach we have sketched is more a vision, rather than an actual research project. However, we strongly advocate the use of high-level concepts from dynamical systems, and mainly attractor landscapes, not just as metaphors but as design guidelines. In addition, we believe that this level of abstraction can provide a common vocabulary and a shared set of categories between researchers in AI and SB, and that this bridge between cell and robot dynamics is worth to be pursued in the future.

14.2 Case study II: Engineering behavioural differentiation in BN controlled robots

14.2.1 Vision

By exploiting the analogy with cell differentiation phenomenology, we want to devise an automatic procedure for the generation of robot controllers able to express *behavioural differentiation*. With *behavioural differentiation*, we refer to the robot ability to express different specialised behaviours triggered by different signals with the further possibility of switchback to unspecialised behaviour(s).

14.2.2 Methodology

We decided to control robot agents via Boolean networks (BNs). This choice has been motivated by the promising results reported in 13.1.1 which have laid the groundwork for the use of BN in the robotic field, and the whole literature, to which this dissertation contributes, which refers to the use of them to reproduce cellular phenomena.

Given the ambitious and challenging objective, we decided to split the problem into two, more tractable, ones:

Subproblem I Development of algorithms for the generation of BNs capable of performing specialised behaviours when used as robot controller. We call this networks *behaviour-BNs*.

Subproblem II Development and engineering of a BN able to express the desired differentiation dynamics and on which the specialised behaviours obtained with the previous point will be mapped. We call this networks *control-BNs*.

In Figure 14.7 we can see a schematic representation of the adopted approach to face the behavioural differentiation problem; moreover, it clarifies the relationship between the two identified subproblems.

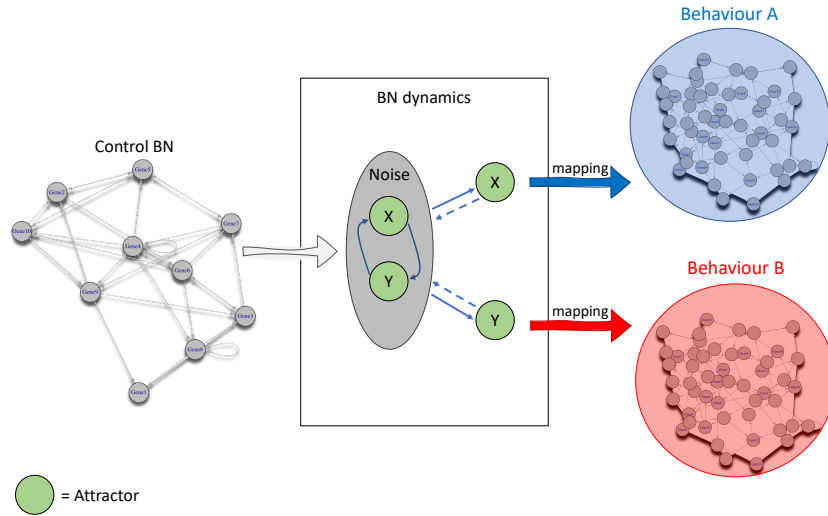


Figure 14.7: This is a schematic representation of the devised approach for reaching robot capable of behavioural differentiation.

After various experimental analyses and improvement phases by trial and error, we came to the formulation of an algorithm for the generation of BN able to satisfy the requirements expressed in the first subproblem. The procedure above is based on the VNS algorithm presented 7.1.2, developed in the

context of reproduction of cellular differentiation phenomenology. The algorithm has successfully produced two BNs capable of expressing *phototaxis* and *anti-phototaxis*, which will represent the specialised behaviours. We refer the reader to the thesis [Cevoli, 2019] for a description of the preliminary results obtained.

We translated the second subproblem into the following requirements. This specific problem formulation has been the result of the necessity of providing a significative but at the same time tractable proof of concept of the proposed vision: Thus, a network that fulfils all the requirements for the second subproblem must present us:

- *two* attractors;
- a dynamics that wander between these two attractors when (intrinsic) noise is present (this can be formally defined by means of the already presented TES_0 concept 6.5),
- while it settles down on a specific attractor upon signal receipt (deterministically associated with the specific destination attractor) and noise disappearance³;
- it must also be able to wander again between the two attractors upon noise re-introduction.

The search for the control-BN has been successfully achieved by means of a *generate-and-test* algorithm, given the modest requirements to be fulfilled.

Lastly, we have linked the three obtained BNs (two *behaviour-BNs* and one *control-BN*) by mapping onto the two attractors of the *control-BN* the two *behaviour-BNs* capable of expressing phototaxis and anti-phototaxis. In this way, the *control-BN* logically determine the robot behaviour but actually the robot behaviour is put in place by the *behaviour-BNs* dynamics. For reasons related to computational costs, we choose among the different specialised behaviour in a random fashion if the network dynamics is in states not belonging to its attractors (transient states).

14.2.3 Results and Conclusion

The Boolean network-based control software has been tested in a simulated environment by means of Webots [Michel, 2004], an open source robot simulator.

The simulated environment dynamics have been subdivided into these phases:

³A similar process can be obtained by thinking to a TES_0 composed by only two attractors that split upon *threshold* increment—and so noise reduction—into two TES s each one composed by one of these attractors.

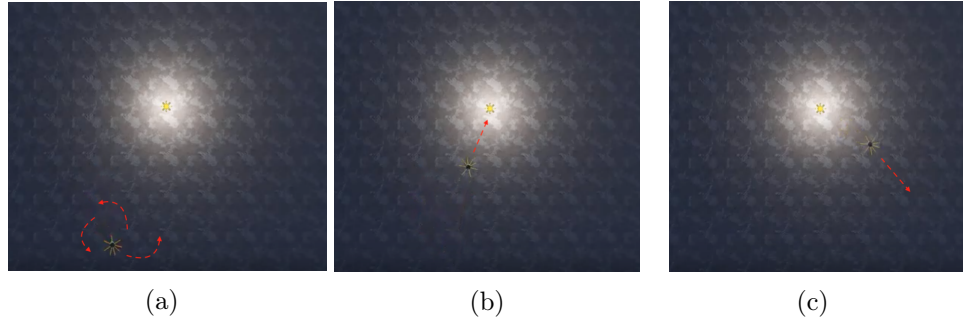


Figure 14.8: a: Noise condition. b: Phototaxis specialised behaviour. c: Anti-phototaxis specialised behaviour. Summary by images of the overall behaviour observed by one of the compositions of networks that have proved to be able to meet the required requirements. The video of the robot behaviour is reported online at this web address <https://www.youtube.com/watch?v=1Tyco4cEFRI&feature=youtu.be>. The experiment, the subject of the video, was carried out by the student Alessandro Cevoli [Cevoli, 2019].

1. noise;
2. the signal that triggers *phototaxis* behaviour;
3. noise;
4. the signal that triggers *anti-phototaxis* behaviour.

The Figure 14.8 shows a sequence of screenshots related to the behaviour of one of the compound networks (two *behaviour-BNs* and one *control-BN*) that met the requirements. This robot behaviour was obtained through the simulated environmental conditions described above.

Much work still needs to be done to achieve the declared vision. In our vision, a single network, in analogy with the genome in biology, can show the complex dynamics observed in this case study. More complex optimisation techniques than those presented here or evolutionary algorithms can represent valid candidates to achieve this goal. But the success of this case study has the quality of demonstrating that robotics can really benefit from the conceptual, analysis and simulation tools developed in the context of the study of cellular differentiation. So, although this is a starting point, it leads us to believe that the direction taken is the right one.

14.3 Case study III: Online adaptiveness or adaptive semantics?

— *Even without Shapirov's coma, we all knew the time would come when a trip through a bloodstream would become necessary.*

We've been planning something like this for a long time and we knew that this skill of mine would be needed.

— *You might have planned an automated crewless ship.*

— *Someday, perhaps, we will, but not yet. We cannot, even now, make the automation equivalent to the versatility and ingenuity of a human brain.* From Isaac Asimov, “Fantastic Voyage II –

Destination Brain”, 1987.

The discussion taking place inside the miniaturised ship of the 1987 Asimov’s novel about the possibility of devising a *crewless ship*, i.e. an autonomous unmanned vehicle, hits one of the main objectives of present-day research agendas in artificial intelligence (AI). In the novel, the argument supporting the choice of miniaturising both ship and human crew is that available autonomous artificial systems are not sufficiently versatile and smart to accomplish a mission inside a human body. Recent technological advances have made it possible to build incredibly small robots, till the size of tens of nanometers. The current smallest robots—built by biological matter—can perform only a few predetermined actions, therefore they can not attain the level of adaptivity and robustness needed for a complex mission. On the other hand, AI software has recently made tremendous advancements and has been proved capable of learning and accomplishing difficult tasks with a high degree of reliability. This software, however, can not be run onto small robots. A viable way for filling this gap is provided by control programs based on unconventional computation, such as the ones derived from cell dynamics models [Roli and Braccini, 2018].

In this chapter we summarise an experiment in BN-robotics⁴ in which the BN controller adapts online to the mission the robot has to accomplish. As the inspiring vision is that of a micro-robot used for medical applications, the adaptation mechanism used in the experiment is minimalistic, so as to facilitate the construction of such a robot for real applications. The key idea is that BNs—mainly critical ones—produce rather complex dynamics [Roli and Braccini, 2018], which can be exploited for producing different kinds of behaviours in the robot.

14.3.1 Experimental setting

In our experiments we used a robot model equipped with 24 proximity sensors (placed evenly along its main circumference) and controlled by two motorised wheels (see Figure 14.9). The robot moves inside a squared arena, delimited by walls, with a central box (see Figure 14.10). The goal we want the robot

⁴Robots whose control software is based on Boolean networks

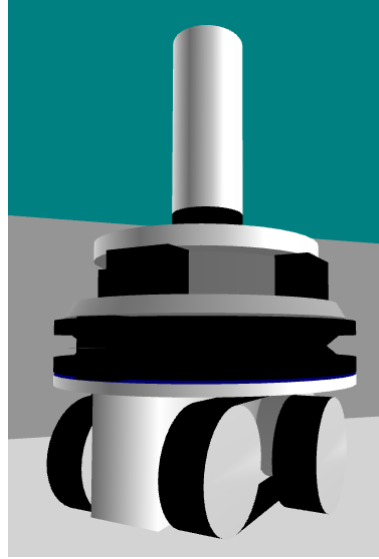


Figure 14.9: The robot used in the experiments.

to achieve is to move as fast as possible around the central box without colliding against walls and the box itself. The robot is controlled by a BN. The coupling between the BN and the robot is as follows: two nodes are randomly chosen and their value is taken to control the two motors. The sensor readings return a value in $[0, 1]$ and so are binarised by a simple step function (with threshold equal to 0.1). The 24 sensors are randomly associated to 24 randomly chosen nodes in the network: at each network update, the binarised values from the sensors are overridden to the current values of the corresponding nodes, so as to provide an external signal to the BN.

The adaptive mechanism is a kind of adaptive random walk. It consists in randomly rewiring up to 6 connections between sensors and BN nodes (excluding output nodes, of course). The robot is then run for 1200 steps (corresponding to 120 seconds of real time, enough for evaluating the robot); if the current binding enables the robot to perform better, then it is kept, otherwise it is rejected and the previous one is taken as the basis for a new perturbation. We remark that the binding between proximity sensors and BN “input” nodes is the only change made to the network: in this way we address the question as to what extent a random BN can indeed provide a sufficient bouquet of behaviours to enable a robot to adapt to a given (minimally cognitive) task.

BNs generated with n nodes, $k = 3$ inputs per node and random Boolean functions defined by means of the bias b .⁵ In the experiments we tested

⁵The bias is the probability of assigning a 1 a truth table entry.

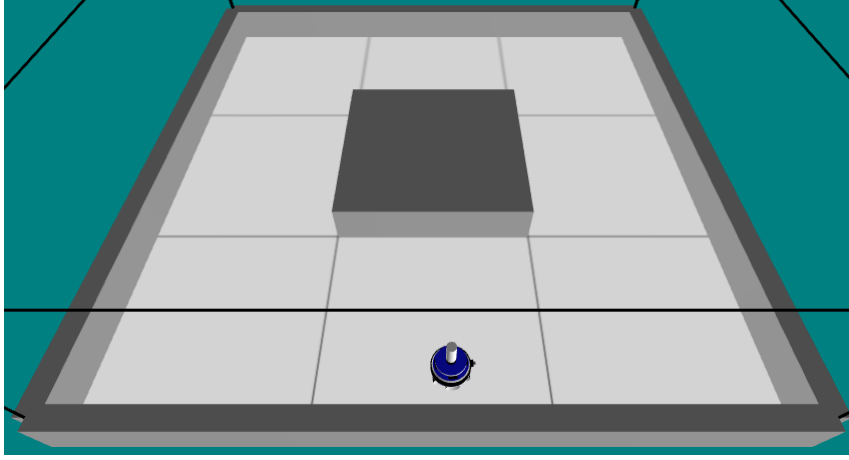


Figure 14.10: The arena used in the experiments.

$n \in \{100, 1000\}$ and $b \in \{0.1, 0.21, 0.5, 0.79, 0.9\}$.⁶

The performance is evaluated by an objective function that is accumulated along the robot execution steps and then normalised. The function is defined as follows:

$$F = (1 - p_{max}) (1 - \sqrt{|b_l - b_r|}) \frac{(b_l + b_r)}{2}$$

where p_{max} is the maximal value returned among the proximity sensors, and b_l and b_r are the binarised values used to control the left and right motor, respectively. The intuition of the function is to favour fast and as much straight as possible trajectories far from the obstacles [Nolfi and Floreano, 2000]. Experiments are run in simulations with ARGoS [Pinciroli et al., 2012].

14.3.2 Results

We run 1000 random replicas for each configuration of BN parameters and collected statistics on the best performance attained after a trial of 1.44×10^4 seconds. As we can observe in Figure 14.11, despite the simple adaptation mechanism, a large fraction of BN attains a good performance.⁷ Notably, critical networks attains the best performance—this result is striking for large BNs ($n = 1000$). This is a further evidence of the conjecture stating that critical systems provide the best trade-off between adaptivity and robustness [Roli et al., 2018]. Observe, however, that just one of the two bias

⁶According to [Luque and Solé, 1997], random BNs with $k = 3$ generated with bias equal to 0.1 or 0.9 are likely to be ordered, with bias equal to 0.5 are likely to be chaotic and bias equal to 0.21 and 0.79 characterises criticality.

⁷Given the evaluation function, for values of F greater than 0.7 the performance is good enough.

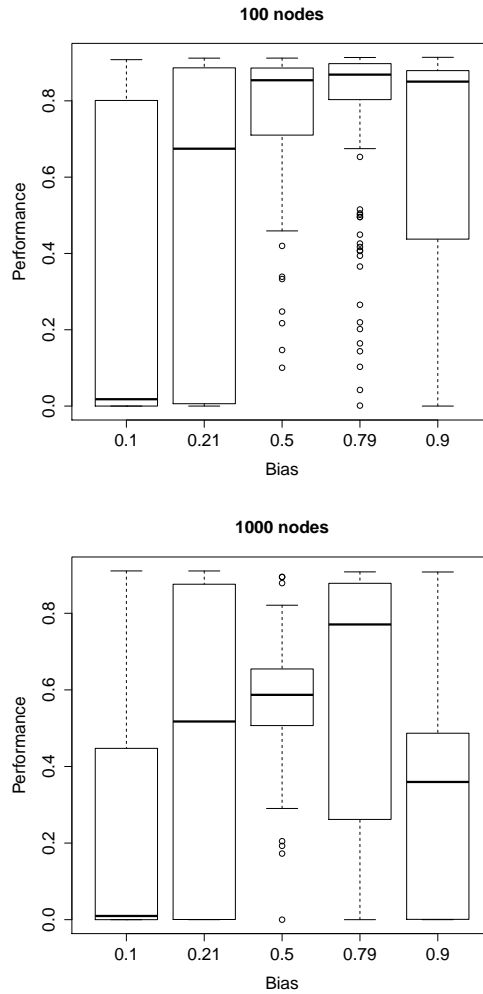


Figure 14.11: Boxplots summarising the performance as a function of BN bias for BNs with $n = 100$ (left) and $n = 1000$ (right).

values corresponding to the critical regime provide good performance. The reason is that in our experiment the symmetry between 0 and 1 is broken, because a 1 means that an obstacle is detected and that motors are on. Indeed, we ran the same experiments with a negative (dual) convention on the values. As expected, results (see Figure 14.12) are perfectly specular to the previous ones.

14.3.3 Conclusion

This experiment represents an attempt to answer some open questions that are also extremely relevant for the origin of life and the evolution of complex

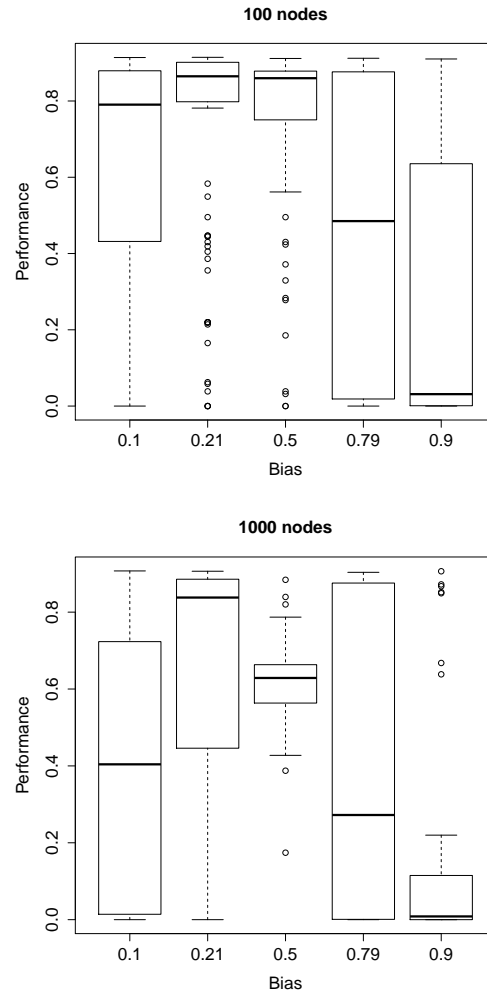


Figure 14.12: Boxplots summarising the performance as a function of BN bias for BNs with $n = 100$ (left) and $n = 1000$ (right) for robots controlled by BNs with a negative encoding (i.e., 0 activate the motor wheels).

organisms. Indeed we believe than the circuits that compose the biological organisms did not appear, during evolution, with a precise and defined semantics. There was no innate “amino acid metabolism pathway” but this firstly has been a working circuit, result of physical constraints and random mutations. Then, the coupling of what we can consider a set of circuits that constituted the embryonic form of those present in today’s cells with the environment has given a meaning to this circuit, a precise meaning caused by the cell system to which it belongs and by the precise coupling of the latter with its specific environment. Its meaning would have most likely been different if the circuit in question had emerged at a different time, organism or

planet. So, we have exploited the above mentioned idea in robotics, trying to verify if online adaptiveness in the form of interpretation of raw material (network) that a robot/organism possesses can be successfully exploited to accomplish a given task. It turned out that networks in the critical dynamic regime were able to accomplish the obstacle avoidance task more successfully, thus providing further evidence regarding the *criticality hypothesis* 4.3. In conclusion, we conjecture that the creation of the semantics of its own circuits, through the environmental coupling and the subsequent refinement of this through the process of evolution itself that acts by strengthening certain mechanisms, can be considered a fundamental step before that which will bring novelties in the organisms. The last mentioned step brings new genetic material into the organism and therefore probably a more suitable substrate on which to build more complex dynamics and functions ⁸.

⁸See the Section 6.4 on “cancer attractors” to grasp the analogy with cell biology evolution counterpart.

Part V

Conclusions and Future
Perspectives

Chapter 15

Ongoing and future work

This dissertation can be somehow considered a starting point of a research direction in the study of differentiation employing more and more realistic and biologically grounded computational models based on Boolean networks. Needless to say, further exploration is needed. Here we propose work in progress and the future one resulting from this research.

A less simplistic noise formulation in BN models for cell differentiation Noise plays a key role in cell processes [Pujadas and Feinberg, 2012, Zwaka, 2006, Huang, 2009], therefore also in cell differentiation. Hence, whatever model we choose (Boolean or not) to represent and simulate cell differentiation dynamics, the influence of noise shouldn't be neglected. In most of the theoretical metaphors or in the dynamical systems view of cell, noise is a fundamental but *abstract actor* in cell dynamics, i.e. a considered but not formally defined aspect. In order to face the problem of the formulation of a computational model for cell differentiation we have to precisely characterise the various noise typologies (intracellular, extracellular, epigenetic, etc.) and their specific implications. In the TES-model 6.5 the differentiation process is strongly correlated with the **intracellular noise level** (modelled by the threshold concept). But the threshold and above all its variation mechanism are externally controlled. In fact the threshold represents an abstraction of the mechanisms implemented by the real cell to control noise. Therefore, it is important to try to identify—guided by experimental data—some autogenous mechanisms, somehow bound to cell's dynamics, through which achieve a noise self-regulation.

An attempt we are currently exploring is to introduce a type of noise determined by the topological regulatory network configuration, this latter determined by the real modelled genome. Indeed, we believe that within a cell the events induced by certain genes (RNAs or proteins production, RNA splicing, etc.) can be considered as simultaneous. Others instead can be perceived by them with variable delays depending on the noise that may

have intervened during the propagation of information between one region and another in the regulatory network. The noise to which we refer with the previous statement can be caused for example by a low number of molecules involved in the information transmission process, aberrant production of nuclear RNA, messenger RNA or protein structures. So we're trying to model this possible phenomenon in Boolean networks.

To put this idea into practice, we have devised a hybrid update scheme: a fraction of nodes are updated in a synchronous way while the remaining fraction is updated in a concurrent way. In detail, with each update of the synchronously updated fraction of nodes, there will be a different update sequence of those updated concurrently. The concurrent fraction is in our vision the modelling counterpart of the unpredictable nature of fine-grain molecular noise that can be found in a biological cell. In Figure 15.1 we can see a partial example of this updating scheme in Boolean models of GRN.

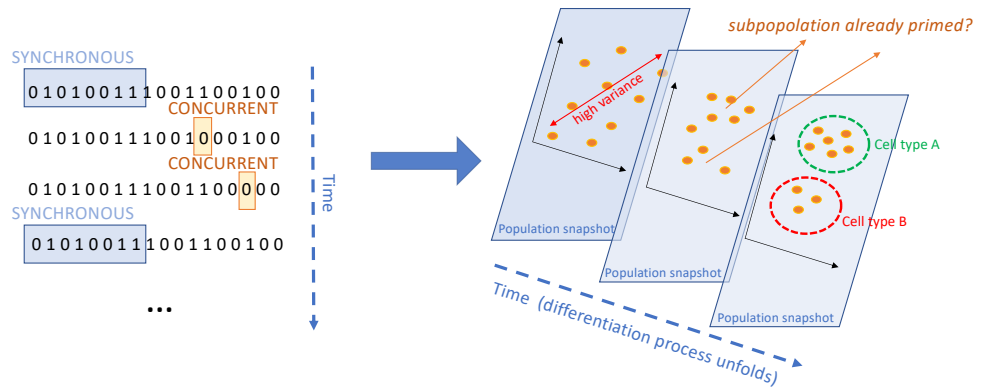


Figure 15.1: This figure is a schematic representation of the hybrid updating scheme proposed with one of its possible phenotypic manifestation in cell populations. In particular, we believe that the employment of this model in populations of Boolean networks can give rise to cells already primed, as observed in recent works, and ultimately to various subpopulations which represent different cell types with no necessity of any external mechanisms.

Preliminary experiments, in progress, are aimed at verifying if this up-

dating scheme can reproduce some phenomena observed during the differentiation process in cell populations. Following the ensemble approach and using random Boolean networks we are now testing whether this model can give rise, in Boolean models, to the same peak of cellular variability observed during the differentiation process just before cell fate decision. The last cited phenomenon is observed with single-cell resolution technologies in two recent noteworthy works [Mojtahedi et al., 2016, Richard et al., 2016].

The experimental attempts we are undertaking are aimed at observing whether with certain fractions of synchronous nodes it is possible to obtain a relative peak of variability in gene expression similar to that reported in the above mentioned works, measured with Shannon entropy as in [Richard et al., 2016], and at the same time check if:

- primed cells—cells enriched for specific lineages [Mojtahedi et al., 2016]¹—exist during the commitment phase;
- there are similarities between the expression vectors of the attractors that represent the cell type(s) before the differentiation step and the reachable cell types after the peak of variability.

Indeed, regarding the second point, we have reasons to believe that a completely asynchronous updating scheme can in principle generate a peak of variability in gene expression similar to those observed in real differentiating populations. Still, the similarities in gene expression among one cell types and the next one typical of differentiation lineages might be lacking.

A Java-based software for BN simulation and analysis During my Ph.D. studies, I have been developing a software tool used to perform the whole set of *in silico* experiments—simulations and analysis—reported in this dissertation. The software—now in a prototype version—is written in Java and is specifically designed for efficient simulation of Boolean network models of genetic regulatory networks. Compared to the already existing software—which most are limited to the simulation of synchronous or asynchronous updating schemes—in our implementations we wanted to raise the abstraction level and support cells dynamics with different updating schemes and different network motifs and topologies (e.g. self-loops), with a support to various biological grounded types of noise—like those above presented.

In the next future, we intend to release it—as done with the software library *diffeRenTES* 7.3—in a stable version under an open-source license. Indeed, we believe that the use of computational tools can be beneficial also for biologists, physicist, and whoever has not a computer science background.

¹Experimentally, different populations of primed cells may be mistakenly believed to be in the same state because they exhibit the same cell surface markers even if they are already directed towards distinct cell fates by means of the expression of other genes [Huang et al., 2009a, Cahan and Daley, 2013].

These tools can be necessary to simulate their hypotheses and to guide the design of *in vitro* experiments. Lastly, it may provide an extremely useful tool also in the process of models refinement: it would be a means to test and formulate new hypotheses in front of the availability of new experimental data on cell dynamics and then drive the formation of new more accurate models, in an iterative process.

Match of model predictions against real data In the era of single-cell technology [Wilson and Göttgens, 2018, Wagner et al., 2018, Farrell et al., 2018, Zheng et al., 2017], we need to check whether Boolean models can be adequate abstractions for the study of cell differentiation and cellular processes in general. We know that Boolean models in principle can provide adequate abstractions for cell differentiation with regard to its generic properties; however, it is necessary to match models previsions against experimental data. To this end, it may be necessary to include further mechanisms into current models or devise new ones. Indeed, in the light of the latest evidence regarding the cell-to-cell variability and the absence of a single “textbook-model” [Pelkmans, 2012] cell, it is necessary to introduce mechanisms that cause this population-level heterogeneity in modelling. For this purpose, new models capable of giving rise to the intrinsic heterogeneity of cell populations, such as the hybrid updating scheme previously outlined, can uncover their origins and causes.

Robotics In the light of the preliminary but promising results obtained using models and abstractions related to the differentiation process in the field of robotics, we intend to investigate them further. Robotics—and swarm robotics in particular given its natural analogy with biological cell populations—represents an invaluable and affordable *testbed* for all the hypotheses, conjectures and implications of the works here presented. Indeed, *wet* or *in vitro* experiments can be really expensive while robots, in simulated or in physical environments, can be the place to combine the whole set of our modelling contributions, contributions which investigate specific—often orthogonal—aspects of the differentiation process.

In addition, for what concerns the behavioural differentiation of robotic agents, behaviours assessment in real-world and the design of a more refined engineering process which can include an *online* adaptation process like the one proposed in Section 14.3 are in our agenda of future works.

Conclusion

The main objective of this thesis is the progress in the understanding of the complex dynamics that give rise to the phenomenology of cell differentiation. The scientific interest on which this analysis is based primarily originates as a consequence of the lack of a *general theoretical framework for cell differentiation*. The approach proposed in this dissertation, based on modelling and dynamical systems theory, had the objective to investigate mechanisms underpinning cell differentiation. This thesis contributes to make a further step towards a formulation of a general theory of cell differentiation.

Although there are issues that require further investigation, this thesis, on the one hand, has made progresses in already (partially) beaten lines of research in complex systems biology, expanding, refining and proposing models of cell differentiation mechanisms, while on the other, it has paved the way for a new research direction in robotics.

In particular, we have verified theoretical predictions of a recently introduced model for cell differentiation by means of stochastic simulations of Boolean networks.

In addition, we proposed a thorough study of the impact of self-loops in random Boolean networks ensembles. Self-loops contribute to the reproduction of differentiation dynamics in modelling. Given their effects on the dynamics of Boolean networks, we also hypothesise that they may have represented evolutionary advantages by providing a better balance between robustness and flexibility.

Moreover, we proposed a new dynamical model of differentiation based on a simplified bio-inspired methylation mechanism in Boolean models of GRNs. This mechanism imposes an arrow of time to the differentiation process by stabilising and limiting the network dynamics. Studies of the combined effects of self-loops and epigenetics—two of the most important mechanisms for controlling differentiation for molecular biologists—have been carried out. Particular combinations of the two mechanisms, when the self-loops have a self-activating effect, bring in BNs not-trivial dynamics, resembling the biological counterparts.

Furthermore, we introduced automatic techniques for the design of Boolean networks capable of expressing the desired differentiation dynamics. Then, these algorithms represented the starting point for the generation of robot

control software in the context of what we called *behavioural differentiation robotics*. Behavioural differentiation robotics is a new branch of research in robotics—proposed by us for the first time—that deals with the designing of robotic agents with the ability to specialise their behaviours in a way similar to that of cells that undergo differentiation. This result is an indirect consequence of our vision, which proposes a paradigm shift in the design of robot behaviours. Indeed, we propose to exploit the conceptual metaphor represented by the *attractor landscape* for the design of robot controllers—but given their generality, it is not limited to them. This idea takes inspiration from the studies conducted in the context of cellular differentiation, but it finds natural application in the design of robotic agents capable of expressing different behaviours.

Finally, we have developed software for the simulation and analysis of Boolean networks, always relating to the reproduction of differentiation phenomena.

The contributions of this work can be considered part of a broader medical research which hopefully in the next future will be able to uncover the origins of differentiation-related diseases. Advancement in drug design and differentiation therapies able to induce cancer cells to switch back towards normal conditions are some of the ambitious objectives of this research line. As we started showing, this research has implications in robotics too. In a futuristic yet feasible scenario, we can imagine micro-robotic agents equipped with differentiation-inspired controller able to perform complex tasks, too risky or impossible for humans. Swarms of agents for fires management, microrobots for space explorations and inoculable nanorobots that can cure ill human organisms by becoming part of them are just some visionary applications of this unconventional bio-inspired computation.

Bibliography

- [Aderem, 2005] Aderem, A. (2005). Systems biology: its practice and challenges. *Cell*, 121(4):511–513.
- [Ahnert and Fink, 2016] Ahnert, S. and Fink, T. (2016). Form and function in gene regulatory networks: the structure of network motifs determines fundamental properties of their dynamical state space. *Journal of the Royal Society Interface*, 13(120):278–289.
- [Albert, 2005] Albert, R. (2005). Scale-free networks in cell biology. *Journal of cell science*, 118(Pt 21):4947–4957.
- [Alberts et al., 2013] Alberts, B., Bray, D., Hopkin, K., Johnson, A. D., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2013). *Essential cell biology*. Garland Science.
- [Aldana, 2003] Aldana, M. (2003). Boolean dynamics of networks with scale-free topology. *Physica D: Nonlinear Phenomena*, 185(1):45 – 66.
- [Aldana et al., 2007] Aldana, M., Balleza, E., Kauffman, S., and Resendiz, O. (2007). Robustness and evolvability in genetic regulatory networks. *Journal of Theoretical Biology*, 245(3):433–448.
- [Aldana et al., 2003] Aldana, M., Coppersmith, S., and Kadanoff, L. P. (2003). Boolean dynamics with random couplings. In *Perspectives and Problems in Nolinear Science*, pages 23–89. Springer.
- [Alon, 2006] Alon, U. (2006). *An introduction to systems biology: design principles of biological circuits*. Chapman and Hall/CRC.
- [Alon, 2007] Alon, U. (2007). Network motifs: theory and experimental approaches. *Nature Reviews Genetics*, 8(6):450.
- [Amé et al., 2006] Amé, J.-M., Halloy, J., Rivault, C., Detrain, C., and Deneubourg, J. L. (2006). Collegial decision making based on social amplification leads to optimal group formation. *Proceedings of the National Academy of Sciences*, 103(15):5835–5840.

- [Astbury, 1961] Astbury, W. T. (1961). Molecular Biology or Ultrastructural Biology? *Nature*, 190(4781):1124.
- [Balakrishnan and Honavar, 1996] Balakrishnan, K. and Honavar, V. (1996). On sensor evolution in robotics. In *Proceedings of the 1st annual conference on genetic programming*, pages 455–460. MIT Press.
- [Bar–Yam, 1997] Bar–Yam, Y. (1997). *Dynamics of Complex Systems*. Studies in nonlinearity. Addison–Wesley, Reading, MA.
- [Barabási et al., 2016] Barabási, A.-L. et al. (2016). *Network science*. Cambridge university press.
- [Bastolla and Parisi, 1997] Bastolla, U. and Parisi, G. (1997). A numerical study of the critical line of kauffman networks. *Journal of Theoretical Biology*, 187(1):117 – 133.
- [Beer, 1995] Beer, R. (1995). A dynamical systems perspective on agent–environment interaction. *Artificial intelligence*, (72):173–215.
- [Beer, 1996] Beer, R. (1996). Toward the evolution of dynamical neural networks for minimally cognitive behavior. *From animals to animats*, 4:421–429.
- [Beer and Williams, 2015] Beer, R. and Williams, P. (2015). Information processing and dynamics in minimally cognitive agents. *Cognitive science*, 39(1):1–38.
- [Beggs and Timme, 2012] Beggs, J. M. and Timme, N. (2012). Being critical of criticality in the brain. *Frontiers in physiology*, 3:163.
- [Benedettini et al., 2014] Benedettini, S., Roli, A., Serra, R., and Villani, M. (2014). Automatic design of Boolean networks for modelling cell differentiation. *Evolution, Complexity and Artificial Life*, 708:77–89.
- [Benedettini et al., 2013] Benedettini, S., Villani, M., Roli, A., Serra, R., Manfroni, M., Gagliardi, A., Pinciroli, C., and Birattari, M. (2013). Dynamical regimes and learning properties of evolved boolean networks. *Neurocomputing*, 99(0):111 – 123.
- [Blum and Roli, 2003] Blum, C. and Roli, A. (2003). Metaheuristics in combinatorial optimization: Overview and conceptual comparison. *ACM Computing Surveys*, 35(3):268–308.
- [Bongard and Pfeifer, 2003] Bongard, J. C. and Pfeifer, R. (2003). Evolving complete agents using artificial ontogeny. In *Morpho-functional Machines: The New Species*, pages 237–258. Springer.

- [Bornholdt and Kauffman, 2019] Bornholdt, S. and Kauffman, S. (2019). Ensembles, dynamics, and cell types: Revisiting the statistical mechanics perspective on cellular regulation. *Journal of theoretical biology*, 467:15–22.
- [Borriello et al., 2018] Borriello, E., Walker, S. I., and Laubichler, M. D. (2018). A unified formal framework for developmental and evolutionary change in gene regulatory network models.
- [Box and Draper, 1987] Box, G. E. and Draper, N. R. (1987). *Empirical model-building and response surfaces*. John Wiley & Sons.
- [Braccini, 2016] Braccini, M. (2016). Automatic design of boolean networks for modelling differentiation trees. Master’s thesis, Università di Bologna. Corso di Studio in Ingegneria e Scienze Informatiche – Cesena.
- [Braccini, 2017] Braccini, M. (2017). Applications of biological cell models in robotics. arXiv:1712.02303. To be submitted.
- [Braccini et al., 2019a] Braccini, M., Montagna, S., and Roli, A. (2019a). Self-loops favour diversification and asymmetric transitions between attractors in boolean network models. In Cagnoni, S., Mordonini, M., Pecori, R., Roli, A., and Villani, M., editors, *Artificial Life and Evolutionary Computation*, pages 30–41, Cham. Springer International Publishing.
- [Braccini et al., 2019b] Braccini, M., Roli, A., Villani, M., Montagna, S., and Serra, R. (2019b). The effects of a simplified model of chromatin dynamics on attractors robustness in random boolean networks with self-loops: an experimental study. In *Proceedings of the XIV International Workshop on Artificial Life and Evolutionary Computation. Rende (CS), Italy, 18-20 September 2019*. To be published.
- [Braccini et al., 2017] Braccini, M., Roli, A., Villani, M., and Serra, R. (2017). *Automatic Design of Boolean Networks for Cell Differentiation*, pages 91–102. Springer International Publishing, Cham.
- [Braccini et al., 2018] Braccini, M., Roli, A., Villani, M., and Serra, R. (2018). A comparison between threshold ergodic sets and stochastic simulation of boolean networks for modelling cell differentiation. In Pelillo, M., Poli, I., Roli, A., Serra, R., Slanzi, D., and Villani, M., editors, *Artificial Life and Evolutionary Computation – 12th Italian Workshop, WIVACE 2017, Revised Selected Papers*, volume 830 of *CCIS*, pages 116–128. Springer.
- [Bull, 2014] Bull, L. (2014). Evolving boolean regulatory networks with epigenetic control. *Biosystems*, 116:36–42.

- [Cahan and Daley, 2013] Cahan, P. and Daley, G. Q. (2013). Origins and implications of pluripotent stem cell variability and heterogeneity. *Nature reviews Molecular cell biology*, 14(6):357.
- [Campbell et al., 2008] Campbell, N. A., Reece, J. B., Urry, L. A., Cain, M. L., Wasserman, S. A., Minorsky, P. V., and Jackson, R. B. (2008). Biology eight edition. *San Francisco: Pearson Education Inc.*
- [Capra and Luisi, 2014] Capra, F. and Luisi, P. L. (2014). *The systems view of life: A unifying vision*. Cambridge University Press.
- [Cariani, 1992] Cariani, P. (1992). Some epistemological implications of devices which construct their own sensors and effectors. In *Toward a practice of autonomous systems: Proceedings of the first European conference on artificial life*, pages 484–493. MIT Press.
- [Cariani, 1993] Cariani, P. (1993). To evolve an ear. Epistemological implications of Gordon Pask’s electrochemical devices. *Systems Research and Behavioral Science*, 10(3):19–33.
- [Cevoli, 2019] Cevoli, A. (2019). Engineering behavioural differentiation in robots controlled by boolean networks. Master’s thesis, Università di Bologna. Corso di Studio in Ingegneria e Scienze Informatiche – Cesena.
- [Chalancon et al., 2012] Chalancon, G., Ravarani, C., Balaji, S., Martinez-Arias, A., Aravind, L., Jothi, R., and Babu, M. (2012). Interplay between gene expression noise and regulatory network architecture. *Trends in genetics*, 28(5):221–232.
- [Chambers et al., 1983] Chambers, J. M., Cleveland, W. S., Kleiner, B., and Tukey, P. A. (1983). *Comparing Data Distributions*, pages 60–63. The Wadsworth Statistics/Probability Series. Wadsworth International Group, Belmont, California.
- [Cohen, 1995] Cohen, P. (1995). *Empirical methods for Artificial Intelligence*. The MIT Press.
- [Crews and Pearson, 2009] Crews, S. T. and Pearson, J. C. (2009). Transcriptional autoregulation in development. *Current Biology*, 19(6):R241–R246.
- [Crutchfield and Young, 1990] Crutchfield, J. and Young, K. (1990). *Computation at the onset of chaos*. In: Complexity, Entropy, and Physics of Information, Addison Wesley.
- [Dawkins, 1976] Dawkins, R. (1976). *The Selfish Gene*. Oxford: Oxford University Press.

- [De Jong, 2002] De Jong, H. (2002). Modeling and simulation of genetic regulatory systems: a literature review. *Journal of computational biology*, 9(1):67–103.
- [Deans and Maggert, 2015] Deans, C. and Maggert, K. A. (2015). What do you mean, “epigenetic”? *Genetics*, 199(4):887–896.
- [Dellaert and Beer, 1994] Dellaert, F. and Beer, R. D. (1994). Co-evolving body and brain in autonomous agents using a developmental model. *Cleveland, OH*, 44106.
- [Dellaert and Beer, 1996] Dellaert, F. and Beer, R. D. (1996). A developmental model for the evolution of complete autonomous agents. *From animals to animats*, 4:393–401.
- [Eggenberger, 1996] Eggenberger, P. (1996). Cell interactions as a control tool of developmental processes for evolutionary robotics. In *Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior*, pages 440–448.
- [Eggenberger, 1997] Eggenberger, P. (1997). Evolving morphologies of simulated 3d organisms based on differential gene expression. In *Proceedings of the fourth european conference on Artificial Life*, pages 205–213.
- [Farrell et al., 2018] Farrell, J. A., Wang, Y., Riesenfeld, S. J., Shekhar, K., Regev, A., and Schier, A. F. (2018). Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. *Science*, 360(6392):eaar3131.
- [Fisher, 2002] Fisher, A. G. (2002). Opinion-decision making in the immune system: cellular identity and lineage choice. *Nature Reviews Immunology*, 2(12):977.
- [Francesca and Birattari, 2016] Francesca, G. and Birattari, M. (2016). Automatic design of robot swarms: achievements and challenges. *Frontiers in Robotics and AI*, 3:29.
- [Francesca et al., 2014] Francesca, G., Brambilla, M., Brutschy, A., Trianni, V., and Birattari, M. (2014). Automode: A novel approach to the automatic design of control software for robot swarms. *Swarm Intelligence*, 8(2):89–112.
- [Furusawa and Kaneko, 2012] Furusawa, C. and Kaneko, K. (2012). A dynamical-systems view of stem cell biology. *Science*, 338:215–217.
- [Garattoni et al., 2013] Garattoni, L., Roli, A., Amaducci, M., Pincioli, C., and Birattari, M. (2013). Boolean network robotics as an intermediate step in the synthesis of finite state machines for robot control. In Liò, P.,

- Miglino, O., Nicosia, G., Nolfi, S., and Pavone, M., editors, *Advances in Artificial Life, ECAL 2013*, pages 372–378. The MIT Press.
- [Geard and Willadsen, 2009] Geard, N. and Willadsen, K. (2009). Dynamical approaches to modeling developmental gene regulatory networks. *Birth Defects Research Part C: Embryo Today: Reviews*, 87(2):131–142.
- [Giere, 2004] Giere, R. N. (2004). How models are used to represent reality. *Philosophy of science*, 71(5):742–752.
- [Gilbert and Barresi, 2016] Gilbert, S. and Barresi, M. (2016). Developmental biology 11th.
- [Gini, 1912] Gini, C. (1912). *Variabilità e mutabilità; contributo allo studio delle distribuzioni e delle relazioni statistiche. [Fasc. I.]*. Tipogr. di P. Cuppini, Bologna. OCLC: 83150327.
- [Graudenzi et al., 2011] Graudenzi, A., Serra, R., Villani, M., Damiani, C., Colacci, A., and Kauffman, S. A. (2011). Dynamical properties of a boolean model of gene regulatory network with memory. *Journal of Computational Biology*, 18(10):1291–303.
- [Guo et al., 2009] Guo, H., Meng, Y., and Jin, Y. (2009). A cellular mechanism for multi-robot construction via evolutionary multi-objective optimization of a gene regulatory network. *BioSystems*, 98(3):193–203.
- [Hansen and Mladenović, 2001] Hansen, P. and Mladenović, N. (2001). Variable neighborhood search: Principles and applications. *European Journal of Operational Research*, 130:449–467.
- [Hardin et al., 2012] Hardin, J., Bertoni, G., and Kleinsmith, L. J. (2012). *Becker’s World of the Cell*. Benjamin Cummings.
- [Harris et al., 2002] Harris, S. E., Sawhill, B. K., Wuensche, A., and Kauffman, S. (2002). A model of transcriptional regulatory networks based on biases in the observed regulation rules. *Complexity*, 7(4):23–40.
- [Helikar et al., 2012] Helikar, T., Kowal, B., McClenathan, S., Bruckner, M., Rowley, T., Madrahimov, A., Wicks, B., Shrestha, M., Limbu, K., and Rogers, J. A. (2012). The cell collective: toward an open and collaborative approach to systems biology. *BMC systems biology*, 6(1):96.
- [Hermesen et al., 2010] Hermesen, R., Ursem, B., and ten Wolde, P. R. (2010). Combinatorial gene regulation using auto-regulation. *PLOS Computational Biology*, 6(6):1–13.
- [Hilborn et al., 2000] Hilborn, R. C. et al. (2000). *Chaos and nonlinear dynamics: an introduction for scientists and engineers*. Oxford University Press on Demand.

- [Hoffmann et al., 2008] Hoffmann, M., Chang, H. H., Huang, S., Ingber, D. E., Loeffler, M., and Galle, J. (2008). Noise-driven stem cell and progenitor population dynamics. *PLOS ONE*, 3(8):1–10.
- [Hotz et al., 2003] Hotz, P. E., Gomez, G., and Pfeifer, R. (2003). Evolving the morphology of a neural network for controlling a foveating retina-and its test on a real robot. In *Artificial Life VIII-8th International Conference on the Simulation and Synthesis of Living Systems*, page 12. Citeseer.
- [Huang et al., 2009a] Huang, A. C., Hu, L., Kauffman, S. A., Zhang, W., and Shmulevich, I. (2009a). Using cell fate attractors to uncover transcriptional regulation of hl60 neutrophil differentiation. *BMC systems biology*, 3(1):20.
- [Huang, 2009] Huang, S. (2009). Non-genetic heterogeneity of cells in development: more than just noise. *Development*, 136(23):3853–3862.
- [Huang, 2012] Huang, S. (2012). The molecular and mathematical basis of waddington’s epigenetic landscape: A framework for post-darwinian biology? *Bioessays*, 34(2):149–157.
- [Huang et al., 2005] Huang, S., Eichler, G., Bar-Yam, Y., and Ingber, D. (2005). Cell fates as high-dimensional attractor states of a complex gene regulatory network. *Physical Review Letters*, 94:128701:1–4.
- [Huang et al., 2009b] Huang, S., Ernberg, I., and Kauffman, S. (2009b). Cancer attractors: a systems view of tumors from a gene network dynamics and developmental perspective. In *Seminars in cell & developmental biology*, volume 20, pages 869–876. Elsevier.
- [Huang et al., 2007] Huang, S., Guo, Y.-P., May, G., and Enver, T. (2007). Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. *Developmental biology*, 305(2):695–713.
- [Huang and Ingber, 2000] Huang, S. and Ingber, D. E. (2000). Shape-dependent control of cell growth, differentiation, and apoptosis: switching between attractors in cell regulatory networks. *Experimental cell research*, 261(1):91–103.
- [Huang and Kauffman, 2013] Huang, S. and Kauffman, S. (2013). How to escape the cancer attractor: rationale and limitations of multi-target drugs. In *Seminars in cancer biology*, volume 23, pages 270–278. Elsevier.
- [Ideker et al., 2001] Ideker, T., Galitski, T., and Hood, L. (2001). A new approach to decoding life: systems biology. *Annual review of genomics and human genetics*, 2(1):343–372.

- [Jakobi, 2003] Jakobi, N. (2003). Harnessing morphogenesis. *On growth, form and computers*, pages 392–404.
- [Jen, 2003] Jen, E. (2003). Stable or robust? what’s the difference? *Complexity*, 8(3):12–18.
- [Jin and Meng, 2011] Jin, Y. and Meng, Y. (2011). Morphogenetic robotics: An emerging new field in developmental robotics. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 41(2):145–160.
- [Johnson et al., 1995] Johnson, N. L., Kotz, S., and Balakrishnan, N. (1995). *Continuous Univariate Distributions*. Wiley.
- [Joo et al., 2018] Joo, J. I., Zhou, J. X., Huang, S., and Cho, K.-H. (2018). Determining relative dynamic stability of cell states using boolean network model. *Scientific reports*, 8(1):12077.
- [Kaneko, 2006] Kaneko, K. (2006). *Life: an introduction to complex systems biology*. Springer.
- [Kaplan and Glass, 2012] Kaplan, D. and Glass, L. (2012). *Understanding nonlinear dynamics*. Springer Science & Business Media.
- [Kauffman, 1969a] Kauffman, S. (1969a). Homeostasis and differentiation in random genetic control networks. *Nature*, 224(5215):177–178.
- [Kauffman, 1969b] Kauffman, S. (1969b). Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of Theoretical Biology*, 22(3):437–467.
- [Kauffman, 1971] Kauffman, S. (1971). Differentiation of malignant to benign cells. *Journal of Theoretical Biology*, 31(3):429–451.
- [Kauffman, 1993] Kauffman, S. (1993). *The Origins of Order: Self-Organization and Selection in Evolution*. Oxford University Press, UK.
- [Kauffman, 1996] Kauffman, S. (1996). *At home in the universe*. Oxford University Press.
- [Kauffman, 2000] Kauffman, S. (2000). *Investigations*. Oxford University Press.
- [Kauffman, 2004] Kauffman, S. (2004). A proposal for using the ensemble approach to understand genetic regulatory networks. *Journal of Theoretical Biology*, 230:581–590.
- [Kauffman and Levin, 1987] Kauffman, S. and Levin, S. (1987). Towards a general theory of adaptive walks on rugged landscapes. *Journal of theoretical Biology*, 128(1):11–45.

- [Kauffman et al., 2003] Kauffman, S., Peterson, C., Samuelsson, B., and Troein, C. (2003). Random boolean network models and the yeast transcriptional network. *Proceedings of the National Academy of Sciences*, 100(25):14796–14799.
- [Kim and Costello, 2017] Kim, M. and Costello, J. (2017). Dna methylation: an epigenetic mark of cellular memory. *Experimental & molecular medicine*, 49(4):e322.
- [Kitano, 2002] Kitano, H. (2002). Systems biology: a brief overview. *science*, 295(5560):1662–1664.
- [Kitano, 2004] Kitano, H. (2004). Biological robustness. *Nature Reviews Genetics*, 5(11):826–837.
- [Kitano et al., 2001] Kitano, H. et al. (2001). *Foundations of systems biology*. MIT press Cambridge.
- [Langton, 1990] Langton, C. (1990). Computation at the edge of chaos: Phase transitions and emergent computation. *Physica D*, 42:12–37.
- [Lodish et al., 2008] Lodish, U. H., Lodish, H., Berk, A., Kaiser, C. A., Kaiser, U. C. A., Krieger, M., Scott, M. P., Bretscher, A., Ploegh, H., Matsudaira, P., Ploegh, U. H., and Matsudaira, U. P. (2008). *Molecular Cell Biology*. W. H. Freeman.
- [Lungarella et al., 2003] Lungarella, M., Metta, G., Pfeifer, R., and Sandini, G. (2003). Developmental robotics: a survey. *Connection Science*, 15(4):151–190.
- [Luque and Solé, 1997] Luque, B. and Solé, R. (1997). Phase transitions in random networks: Simple analytic determination of critical points. *Physical Review E*, 55(1).
- [M Biava et al., 2011] M Biava, P., Basevi, M., Biggiero, L., Borgonovo, A., Borgonovo, E., and Burigana, F. (2011). Cancer cell reprogramming: stem cell differentiation stage factors and an agent based model to optimize cancer treatment. *Current pharmaceutical biotechnology*, 12(2):231–242.
- [McAdams and Arkin, 1997] McAdams, H. and Arkin, A. (1997). Stochastic mechanisms in gene expression. *Proceedings of the National Academy of Sciences*, 94(3):814–819.
- [Meng and Guo, 2012] Meng, Y. and Guo, H. (2012). Evolving network motifs based morphogenetic approach for self-organizing robotic swarms. In *Proceedings of the 14th annual conference on Genetic and evolutionary computation*, pages 137–144. ACM.

- [Meng et al., 2013] Meng, Y., Guo, H., and Jin, Y. (2013). A morphogenetic approach to flexible and robust shape formation for swarm robotic systems. *Robotics and Autonomous Systems*, 61(1):25–38.
- [Meng et al., 2010] Meng, Y., Zheng, Y., and Jin, Y. (2010). A morphogenetic approach to self-reconfigurable modular robots using a hybrid hierarchical gene regulatory network. In *Proc. of the Alife XII Conference, Odense, Denmark, 2010*, pages 765–772.
- [Michel, 2004] Michel, O. (2004). Webotstm: Professional mobile robot simulation. *CoRR*, abs/cs/0412052.
- [Michel and Biondi, 1995a] Michel, O. and Biondi, J. (1995a). From the chromosome to the neural network. In *Artificial Neural Nets and Genetic Algorithms*, pages 80–83. Springer.
- [Michel and Biondi, 1995b] Michel, O. and Biondi, J. (1995b). Morphogenesis of neural networks. *Neural Processing Letters*, 2(1):9–12.
- [Milano and Roli, 2000] Milano, M. and Roli, A. (2000). Solving the Satisfiability problem through Boolean networks. In *AI*IA99: Advances in artificial intelligence*, volume 1792 of *lncs*, pages 72–83. springer.
- [Milo et al., 2009] Milo, R., Jorgensen, P., Moran, U., Weber, G., and Springer, M. (2009). Bionumbers—the database of key numbers in molecular and cell biology. *Nucleic acids research*, 38(suppl_1):D750–D753.
- [Milo et al., 2002] Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. (2002). Network motifs: simple building blocks of complex networks. *Science*, 298(5594):824–827.
- [Miyamoto et al., 2015] Miyamoto, T., Furusawa, C., and Kaneko, K. (2015). Pluripotency, differentiation, and reprogramming: a gene expression dynamics model with epigenetic feedback regulation. *PLoS computational biology*, 11(8):e1004476.
- [Mojtahedi et al., 2016] Mojtahedi, M. et al. (2016). Cell fate decision as high-dimensional critical state transition. *PLOS Biology*, 14(12):e2000640:1–28.
- [Montagna et al., 2018] Montagna, S., Braccini, M., and Roli, A. (2018). The impact of self-loops in random boolean network dynamics: A simulation analysis. In Pelillo, M., Poli, I., Roli, A., Serra, R., Slanzi, D., and Villani, M., editors, *Artificial Life and Evolutionary Computation – 12th Italian Workshop, WIVACE 2017, Revised Selected Papers*, volume 830 of *CCIS*, pages 104–115. Springer.

- [Montagna et al., 2020] Montagna, S., Braccini, M., and Roli, A. (2020). The impact of self-loops on boolean networks attractor landscape and implications for cell differentiation modelling. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*. To be published.
- [Montagna et al., 2010a] Montagna, S., Donati, N., and Omicini, A. (2010a). An agent-based model for the pattern formation in drosophila melanogaster. In *ALIFE*, pages 110–117.
- [Montagna and Omicini, 2017] Montagna, S. and Omicini, A. (2017). Agent-based modelling in multicellular systems biology. In *Multi-Agent-Based Simulations Applied to Biological and Environmental Systems*, pages 159–178. IGI Global.
- [Montagna et al., 2010b] Montagna, S., Omicini, A., and Ricci, A. (2010b). A multiscale agent-based model of morphogenesis in biological systems. In *WOA*. Citeseer.
- [Montagna et al., 2015] Montagna, S., Viroli, M., and Roli, A. (2015). A framework supporting multi-compartment stochastic simulation and parameter optimisation for investigating biological system development. *SIMULATION: Transactions of the Society for Modeling and Simulation International*.
- [Moris et al., 2016] Moris, N., Pina, C., and Arias, A. M. (2016). Transition states and cell fate decisions in epigenetic landscapes. *Nature Reviews Genetics*, 17(11):693.
- [Müssel et al., 2010] Müssel, C., Hopfensitz, M., and Kestler, H. A. (2010). Boolnet – an r package for generation, reconstruction and analysis of boolean networks. *Bioinformatics*, 26(10):1378–1380.
- [Nicolis and Prigogine, 1989] Nicolis, G. and Prigogine, I. (1989). *Exploring complexity: an introduction*. W.H. Freeman, New York.
- [Nolfi and Floreano, 2000] Nolfi, S. and Floreano, D. (2000). *Evolutionary robotics*. The MIT Press, Cambridge, MA.
- [Nykter et al., 2008] Nykter, M., Price, N., Aldana, M., Ramsey, S., Kauffman, S., Hood, L., Yli-Harja, O., and Shmulevich, I. (2008). Gene expression dynamics in the macrophage exhibit criticality. In *Proceedings of the National Academy of Sciences, USA*, volume 105, pages 1897–1900.
- [Oh and Jin, 2014] Oh, H. and Jin, Y. (2014). Evolving hierarchical gene regulatory networks for morphogenetic pattern formation of swarm robots. In *Evolutionary Computation (CEC), 2014 IEEE Congress on*, pages 776–783. IEEE.

- [Olsson et al., 2006] Olsson, L., Nehaniv, C., and Polani, D. (2006). From unknown sensors and actuators to actions grounded in sensorimotor perceptions. *Connection Science*, 18(2):121–144.
- [Packard, 1988] Packard, N. (1988). Adaptation toward the edge of chaos. In *Dynamic Patterns in Complex Systems*, pages 293–301.
- [Paroni et al., 2016] Paroni, A., Graudenzi, A., Caravagna, G., Damiani, C., Mauri, G., and Antoniotti, M. (2016). Cabernet: a cytoscape app for augmented boolean models of gene regulatory networks. *BMC bioinformatics*, 17(1):64.
- [Peláez et al., 2015] Peláez, N., Gavalda-Miralles, A., Wang, B., Navarro, H. T., Gudjonson, H., Rebay, I., Dinner, A. R., Katsaggelos, A. K., Amaral, L. A., and Carthew, R. W. (2015). Dynamics and heterogeneity of a fate determinant during transition towards cell differentiation. *Elife*, 4:e08924.
- [Pelkmans, 2012] Pelkmans, L. (2012). Using cell-to-cell variability—a new era in molecular biology. *Science*, 336(6080):425–426.
- [Perino and Veenstra, 2016] Perino, M. and Veenstra, G. (2016). Chromatin control of developmental dynamics and plasticity. *Developmental Cell*, 38(6):610–620.
- [Pfeifer and Bongard, 2006] Pfeifer, R. and Bongard, J. (2006). *How the Body Shapes the Way We Think: A New View of Intelligence*. MIT Press, Cambridge, MA.
- [Pfeifer and Scheier, 1999] Pfeifer, R. and Scheier, C. (1999). *Understanding intelligence*. MIT Press.
- [Pinciroli et al., 2012] Pinciroli, C., Trianni, V., O’Grady, R., Pini, G., Brutschy, A., Brambilla, M., Mathews, N., Ferrante, E., Di Caro, G., Ducatelle, F., Birattari, M., Gambardella, L., and Dorigo, M. (2012). AR-GoS: a modular, multi-engine simulator for heterogeneous swarm robotics. *Swarm Intelligence*, 6(4):271–295.
- [Pinho et al., 2014] Pinho, R., Garcia, V., Irimia, M., and Feldman, M. W. (2014). Stability depends on positive autoregulation in boolean gene regulatory networks. *PLOS Computational Biology*, 10(11):1–14.
- [Piscopo, 2013] Piscopo, C. (2013). *The metaphysical nature of the non-adequacy claim: an epistemological analysis of the debate on probability in artificial intelligence*. Springer.
- [Prokopenko, 2013] Prokopenko, M. (2013). Information dynamics at the edge of chaos: Measures, examples, and principles. In *Proceedings of IEEE Symposium of Artificial Life - ALIFE 2013*.

- [Pujadas and Feinberg, 2012] Pujadas, E. and Feinberg, A. P. (2012). Regulated noise in the epigenetic landscape of development and disease. *Cell*, 148(6):1123–1131.
- [Quick et al., 2003] Quick, T., Nehaniv, C. L., Dautenhahn, K., and Roberts, G. (2003). Evolving embodied genetic regulatory network-driven control systems. In *Advances in Artificial Life*, pages 266–277. Springer.
- [Raj et al., 2010] Raj, A., Rifkin, S., Andersen, E., and Van Oudenaarden, A. (2010). Variability in gene expression underlies incomplete penetrance. *Nature*, 463(7283):913–918.
- [Ribeiro and Kauffman, 2007] Ribeiro, A. and Kauffman, S. (2007). Noisy attractors and ergodic sets in models of gene regulatory networks. *Journal of theoretical biology*, 247(4):743–755.
- [Richard et al., 2016] Richard, A., Boullu, L., Herbach, U., Bonnafoux, A., Morin, V., Vallin, E., Guillemin, A., Gao, N. P., Gunawan, R., Cosette, J., et al. (2016). Single-cell-based analysis highlights a surge in cell-to-cell molecular variability preceding irreversible commitment in a differentiation process. *PLoS biology*, 14(12):e1002585.
- [Roli et al., 2012] Roli, A., Benedettini, S., Birattari, M., Pinciroli, C., Serra, R., and Villani, M. (2012). A preliminary study on bn-robots’ dynamics. In *Proceedings of the Italian Workshop on Artificial Life and Evolutionary Computation (WIVACE 2012)*, pages 1–4.
- [Roli and Braccini, 2018] Roli, A. and Braccini, M. (2018). Attractor landscape: A bridge between robotics and synthetic biology. *Complex Systems*, 27:229–248.
- [Roli et al., 2011a] Roli, A., Manfroni, M., Pinciroli, C., and Birattari, M. (2011a). On the design of Boolean network robots. In Di Chio, C., Cagnoni, S., Cotta, C., Ebner, M., Ekárt, A., Esparcia-Alcázar, A., Merelo, J., Neri, F., Preuss, M., Richter, H., Togelius, J., and Yannakakis, G., editors, *Applications of Evolutionary Computation*, volume 6624 of *lncs*, pages 43–52. Springer, Heidelberg, Germany.
- [Roli et al., 2011b] Roli, A., Manfroni, M., Pinciroli, C., and Birattari, M. (2011b). On the design of boolean network robots. In *Applications of Evolutionary Computation*, pages 43–52. Springer.
- [Roli et al., 2018] Roli, A., Villani, M., Filisetti, A., and Serra, R. (2018). Dynamical criticality: overview and open questions. *Journal of Systems Science and Complexity*, 31(3):647–663.

- [Roli et al., 2015] Roli, A., Villani, M., Serra, R., Benedettini, S., Pincirolì, C., and Birattari, M. (2015). Dynamical properties of artificially evolved boolean network robots. In *AI*IA2015: Advances in artificial intelligence*, volume 9336 of *lncs*, pages 45–57. springer.
- [Roli et al., 2013] Roli, A., Villani, M., Serra, R., Garattoni, L., Pincirolì, C., and Birattari, M. (2013). Identification of dynamical structures in artificial brains: An analysis of boolean network controlled robots. In *AI*IA2013: Advances in artificial intelligence*, volume 8249 of *lncs*, pages 324–335. springer.
- [Rosenblueth and Wiener, 1945] Rosenblueth, A. and Wiener, N. (1945). The role of models in science. *Philosophy of science*, 12(4):316–321.
- [Russell and Norvig, 2009] Russell, S. and Norvig, P. (2009). *Artificial Intelligence: A Modern Approach*. Prentice Hall, Upper Saddle River, NJ, 3rd edition.
- [Sayama, 2015] Sayama, H. (2015). *Introduction to the modeling and analysis of complex systems*. Open SUNY Textbooks.
- [Schuettengruber and Cavalli, 2009] Schuettengruber, B. and Cavalli, G. (2009). Recruitment of polycomb group complexes and their role in the dynamic regulation of cell fate choice. *Development*, 136(21):3531–3542.
- [Serra and Villani, 2006] Serra, R. and Villani, M. (2006). Agents, equations and all that: on the role of agents in understanding complex systems. In *Reasoning, Action and Interaction in AI Theories and Systems*, pages 159–175. Springer.
- [Serra et al., 2004a] Serra, R., Villani, M., and Agostini, L. (2004a). On the dynamics of random boolean networks with scale-free outgoing connections. *Physica A: Statistical Mechanics and its Applications*, 339(3-4):665 – 673.
- [Serra et al., 2010] Serra, R., Villani, M., Barbieri, A., Kauffman, S., and Colacci, A. (2010). On the dynamics of random Boolean networks subject to noise: attractors, ergodic sets and cell types. *Journal of Theoretical Biology*, 265(2):185–93.
- [Serra et al., 2008] Serra, R., Villani, M., Graudenzi, A., Colacci, A., and Kauffman, S. A. (2008). The simulation of gene knock-out in scale-free random boolean models of genetic networks. *Networks and heterogenous media*, 2(3):333–343.
- [Serra et al., 2007] Serra, R., Villani, M., Graudenzi, A., and Kauffman, S. (2007). Why a simple model of genetic regulatory networks describes the

- distribution of avalanches in gene expression data. *Journal of Theoretical Biology*, 246(3):449 – 460.
- [Serra et al., 2006] Serra, R., Villani, M., Graudenzi, A., and Kauffman, S. A. (2006). On the distribution of small avalanches in random boolean networks. In Ruusovori, P. and et al, editors, *Proceedings of the 4th TICSP workshop on computational systems biology*, pages 93–96. Tampere: Juvenes print.
- [Serra et al., 2004b] Serra, R., Villani, M., and Semeria, A. (2004b). Genetic network models and statistical properties of gene expression data in knock-out experiments. *Journal of Theoretical Biology*, 227:149–157.
- [Serra and Zanarini, 1990] Serra, R. and Zanarini, G. (1990). *Complex Systems and Cognitive Processes*. Springer.
- [Shen-Orr et al., 2002] Shen-Orr, S. S., Milo, R., Mangan, S., and Alon, U. (2002). Network motifs in the transcriptional regulation network of escherichia coli. *Nature Genetics*, 31(1):64–68.
- [Shetty et al., 2008] Shetty, R. P., Endy, D., and Knight, T. F. (2008). Engineering biobrick vectors from biobrick parts. *Journal of Biological Engineering*, 2(1):5.
- [Shmulevich et al., 2005] Shmulevich, I., Kauffman, S., and Aldana, M. (2005). Eukaryotic cells are dynamically ordered or critical but not chaotic. *Proceedings of the National Academy of Sciences*, 102(38):13439–13444.
- [Stevens, 2004] Stevens, C. F. (2004). Systems biology versus molecular biology. *Current Biology*, 14(2):R51–R52.
- [Strogatz, 2018] Strogatz, S. H. (2018). *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering*. CRC Press.
- [Su et al., 2017] Su, H., Wang, G., Yuan, R., Wang, J., Tang, Y., Ao, P., and Zhu, X. (2017). Decoding early myelopoiesis from dynamics of core endogenous network. *Science China Life Sciences*, 60(6):627–646.
- [Thomas, 2002] Thomas, R. (2002). Laws for the dynamics of regulatory networks. *International Journal of Developmental Biology*, 42(3):479–485.
- [Thomas et al., 1976] Thomas, R., Gathoye, A.-M., and Lambert, L. (1976). A complex control circuit: Regulation of immunity in temperate bacteriophages. *European Journal of Biochemistry*, 71(1):211–227.
- [Thomas et al., 1995] Thomas, R., Thieffry, D., and Kaufman, M. (1995). Dynamical behaviour of biological regulatory networks—i. biological role of feedback loops and practical use of the concept of the loop-characteristic state. *Bulletin of mathematical biology*, 57(2):247–276.

- [Turner et al., 2017] Turner, A. P., Caves, L. S., Stepney, S., Tyrrell, A. M., and Lones, M. A. (2017). Artificial epigenetic networks: automatic decomposition of dynamical control tasks using topological self-modification. *IEEE transactions on neural networks and learning systems*, 28(1):218–230.
- [Turner et al., 2013] Turner, A. P., Lones, M. A., Fuente, L. A., Stepney, S., Caves, L. S., and Tyrrell, A. M. (2013). The incorporation of epigenetics in artificial gene regulatory networks. *BioSystems*, 112(2):56–62.
- [Villani et al., 2011] Villani, M., Barbieri, A., and Serra, R. (2011). A dynamical model of genetic networks for cell differentiation. *PloS one*, 6(3):e17703.
- [Villani et al., 2017] Villani, M., Campioli, D., Damiani, C., Roli, A., Filisetti, A., and Serra, R. (2017). Dynamical regimes in non-ergodic random boolean networks. *Natural Computing*, 16(2):353–363.
- [Villani et al., 2018] Villani, M., La Rocca, L., Kauffman, S., and Serra, R. (2018). Dynamical criticality in gene regulatory networks. *Complexity*, 2018.
- [Villani and Serra, 2013a] Villani, M. and Serra, R. (2013a). On the dynamical properties of a model of cell differentiation. *EURASIP Journal on Bioinformatics and Systems Biology*, 2013(1):4.
- [Villani and Serra, 2013b] Villani, M. and Serra, R. (2013b). On the dynamical properties of a model of cell differentiation. *EURASIP Journal on Bioinformatics and Systems Biology*, 4:1–8.
- [Von Neumann et al., 1966] Von Neumann, J., Burks, A. W., et al. (1966). Theory of self-reproducing automata. *IEEE Transactions on Neural Networks*, 5(1):3–14.
- [Waddington, 1942] Waddington, C. H. (1942). *The epigenotype*. Endeavour.
- [Waddington, 1957] Waddington, C. H. (1957). *The strategy of the genes*. Allen and Unwin, London.
- [Wagner et al., 2018] Wagner, D. E., Weinreb, C., Collins, Z. M., Briggs, J. A., Megason, S. G., and Klein, A. M. (2018). Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science*, 360(6392):981–987.
- [Wang et al., 2015] Wang, Z., Butner, J. D., Kerketta, R., Cristini, V., and Deisboeck, T. S. (2015). Simulating cancer growth with multiscale agent-based modeling. In *Seminars in cancer biology*, volume 30, pages 70–78. Elsevier.

- [Wilson and Göttgens, 2018] Wilson, N. K. and Göttgens, B. (2018). Single-cell sequencing in normal and malignant hematopoiesis. *HemaSphere*, 2(2):e34.
- [Wuensche, 1996] Wuensche, A. (1996). Discrete dynamics lab (ddlab). *Software and Manual*.
- [Yeger-Lotem et al., 2004] Yeger-Lotem, E., Sattath, S., Kashtan, N., Itzkovitz, S., Milo, R., Pinter, R. Y., Alon, U., and Margalit, H. (2004). Network motifs in integrated cellular networks of transcription–regulation and protein–protein interaction. *Proceedings of the National Academy of Sciences of the United States of America*, 101(16):5934–5939.
- [Yuan et al., 2016] Yuan, R., Zhu, X., Radich, J. P., and Ao, P. (2016). From molecular interaction to acute promyelocytic leukemia: Calculating leukemogenesis and remission from endogenous molecular-cellular network. *Scientific reports*, 6:24307.
- [Zenil et al., 2018] Zenil, H., Kiani, N., Marabita, F., Deng, Y., Elias, S., Schmidt, A., Ball, G., and Tegner, J. (2018). An algorithmic information calculus for causal discovery and reprogramming systems. *bioRxiv*.
- [Zhang et al., 2009] Zhang, L., Wang, Z., Sagotsky, J. A., and Deisboeck, T. S. (2009). Multiscale agent-based cancer modeling. *Journal of mathematical biology*, 58(4-5):545–559.
- [Zheng et al., 2017] Zheng, C., Zheng, L., Yoo, J.-K., Guo, H., Zhang, Y., Guo, X., Kang, B., Hu, R., Huang, J. Y., Zhang, Q., et al. (2017). Landscape of infiltrating t cells in liver cancer revealed by single-cell sequencing. *Cell*, 169(7):1342–1356.
- [Zhou et al., 2016] Zhou, J., Samal, A., Fouquier d’Hérouël, A., Price, N., and Huang, S. (2016). Relative stability of network states in boolean network models of gene regulation in development. *Biosystems*, 142-143:15–24.
- [Zhou et al., 2014] Zhou, J. X., Qiu, X., Fouquier d’Hérouël, A., and Huang, S. (2014). Discrete gene network models for understanding multicellularity and cell reprogramming: from network structure to attractor landscape. *Computational Systems Biology*, pages 241–276.
- [Ziemke, 2005] Ziemke, T. (2005). Cybernetics and embodied cognition: on the construction of realities in organisms and robots. *Kybernetes*, 34(1/2):118–128.
- [Zlochin et al., 2004] Zlochin, M., Birattari, M., Meuleau, N., and Dorigo, M. (2004). Model-based search for combinatorial optimization: A critical survey. *Annals of Operations Research*, 131(1):373–395.

- [Zwaka, 2006] Zwaka, T. P. (2006). Keeping the noise down in es cells. *Cell*, 127(7):1301–1302.

List of Figures

1	Concept map concerning the relationships between the original contributions produced by the thesis work.	2
1.1	A schematic representation of a cell, DNA, chromosomes and genes.	12
1.2	Example of differentiation tree of hematopoiesis.	17
2.1	Depiction of the systems biology approach.	23
2.2	Organisational map of complex systems science.	25
3.1	Modelling approaches.	31
5.1	Example of a Boolean network.	42
5.2	Example of attractors representation in DDlab.	42
5.3	Phase transitions in RBN.	44
6.1	Epigenetic landscape and differentiation.	49
7.1	Differentiation trees used in the search process.	62
7.2	Boxplot summarising the results for differentiation trees 1 to 4.	64
7.3	Boxplot summarising the results for differentiation trees 5 to 9.	65
7.4	An example of a TES-tree.	69
7.5	Incompatibilities in the case of one flip every 1, 5, 10, 15, 20 steps.	71
7.6	Incompatibilities in the case of one flip every 50, 100, 200, 500 steps.	72
7.7	Incompatibilities in the case of one flip every 1000, 5000, 10000, 20000, 50000 steps.	73
7.8	Distribution of the median no. of noise events during transients in stories.	74
7.9	Detail of Figure 7.8 on logarithmic scale.	74
7.10	Distribution of the median no. of multiple flips in attractors of stories.	75
7.11	Detail of Figure 7.10 on logarithmic scale.	75
8.1	Self-loops addition modes.	82

8.2	Average no. of attractors as a function of the fraction of self-loops added with RND functions.	84
8.3	Average no. of attractors as a function of the fraction of self-loops added with OR function.	85
8.4	Average no. of attractors as a function of the fraction of self-loops with RND function, introduced by rewiring.	85
8.5	Average no. of attractors as a function of the fraction of self-loops with OR function, introduced by rewiring.	86
8.6	Average no. of attractors as a function of the no. of nodes with an OR function.	87
8.7	Statistics of f_1 in the case with $k = 2$ and self-loops combined in OR.	93
8.8	Statistics of f_1 in the case with self-loops added to the nodes in OR.	94
8.9	Statistics of f_1 in the case with $k = 2$ and self-loops combined as a random function.	94
8.10	Statistics of f_1 in the case with self-loops added to the nodes in a random function.	95
8.11	Statistics of f_2 in the case with $k = 2$ and self-loops combined in OR.	95
8.12	Statistics of f_2 in the case with self-loops added to the nodes in OR.	96
8.13	Statistics of f_2 in the case with $k = 2$ and self-loops combined as a random function.	96
8.14	Statistics of f_2 in the case with self-loops added to the nodes in a random function.	97
8.15	Average no. of attractors in RBNs with 15 nodes as a fraction of self-loop varying from 0 to 1.	101
8.16	Comparison between the mathematical model and experimental results.	104
8.17	Density function of the no. of attractors for CONST-OR case.	106
8.18	ECDF of the no. of attractors for CONST-OR case.	107
8.19	δ_{min} and δ_{max} as a function of the no. of attractors for CONST-OR case.	109
8.20	δ_{min} , δ_{max} and ATM diagonal values as a function of the no. of attractors for CONST-OR case.	112
8.21	δ_{min} and δ_{max} as a function of the fraction of nodes with a self-loop for CONST-OR case.	112
8.22	Distribution of the fraction of nodes with a self-loop in BNs from Cell Collective.	115
10.1	Representation of the methylation model.	122
10.2	Distribution of the no. of attractors as the no. of frozen nodes increases.	124

10.3	Distribution of the no. of fixed points over the number of attractors as the no. of frozen nodes increases.	125
10.4	Distribution of the Derrida parameters as the no. of frozen nodes increases.	126
10.5	The trend of the no. of equal reached attractors considered in pair.	127
10.6	The trend of the no. of <i>diversities</i> , triplets randomly chosen among non-already frozen nodes.	128
10.7	The trend of the no. of <i>diversities</i> , triplets randomly chosen among nodes with value 1.	129
10.8	Mean attractor number in BNs with self-loops and frozen nodes. Self-loops added in OR.	133
10.9	Mean fraction of fixed points in BNs with self-loops and frozen nodes. Self-loops added in OR.	134
10.10	Mean attractor number in BNs with self-loops and frozen nodes. Self-loops added in AND.	135
10.11	Mean fraction of fixed points in BNs with self-loops and frozen nodes. Self-loops added in AND.	135
10.12	Mean attractor number in BNs with self-loops and frozen nodes. Self-loops added with RND function.	136
10.13	Mean fraction of fixed points in BNs with self-loops and frozen nodes. Self-loops added with RND function.	136
10.14	Mean attractor robustness in BNs with self-loops and frozen nodes. Self-loops added in OR.	137
10.15	Mean attractor robustness in BNs with self-loops and frozen nodes. Self-loops added in AND.	138
10.16	Mean attractor robustness in BNs with self-loops and frozen nodes. Self-loops added with RND function.	138
13.1	Examples of evolved forms by means of the AES.	156
13.2	An example of evolved agent for block pushing.	160
13.3	Morphogenetic and Epigenetic Robotics.	161
13.4	Conceptual diagram for coevolving the development of arm and control.	163
14.1	State graph of the network defined in Figure 5.1.	166
14.2	Attractor graph of the network defined in Figure 5.1.	167
14.3	Main phases of the phototaxis behaviour of a group of robots.	169
14.4	Main phases of the phototaxis behaviour of a group of robots. Dynamic equilibrium case.	170
14.5	Corridor scenario: the robot points towards a goal and its movement triggered by an external stimulus.	171
14.6	State graph of the network controlling the robot.	172
14.7	Schematic representation for behavioural differentiation in robots.	174

14.8	Summary of the attained behavioural differentiation.	176
14.9	The robot used in the experiments.	178
14.10	The arena used in the experiments.	179
14.11	Boxplots summarising the performance as a function of BN bias.	180
14.12	Boxplots summarising the performance as a function of BN bias; negative encoding case.	181
15.1	Hybrid updating scheme.	186
A.1	Density function of the no. of attractors for AUGM-OR case.	217
A.2	ECDF of the no. of attractors for AUGM-OR case.	218
A.3	δ_{min} and δ_{max} as a function of the no. of attractors for AUGM-OR case.	219
A.4	δ_{min} , δ_{max} and ATM diagonal values as a function of the no. of attractors for AUGM-OR case	220
A.5	δ_{min} and δ_{max} as a function of the fraction of nodes with a self-loop for AUGM-OR case.	220
A.6	Density function of the no. of attractors for CONST-RND case.	221
A.7	ECDF of the no. of attractors for CONST-RND case.	221
A.8	δ_{min} and δ_{max} as a function of the no. of attractors for CONST-RND case.	222
A.9	δ_{min} , δ_{max} and ATM diagonal values as a function of the no. of attractors for CONST-RND case.	223
A.10	δ_{min} and δ_{max} as a function of the fraction of nodes with a self-loop for CONST-RND case.	223
A.11	Density function of the no. of attractors for AUGM-RND case.	224
A.12	ECDF of the number of attractors for AUGM-RND case.	225
A.13	δ_{min} and δ_{max} as a function of the no. of attractors for AUGM-RND case.	226
A.14	δ_{min} , δ_{max} and ATM diagonal values as a function of the no. of attractors for AUGM-RND case.	227
A.15	δ_{min} and δ_{max} as a function of the fraction of nodes with a self-loop for AUGM-RND case.	227
A.16	Analytical model results for AUGM-RND and CONST-RND cases.	228
A.17	Analytical model results for the CONST-OR case.	228
A.18	Analytical model results for the AUGM-OR case.	229

List of Tables

8.1	Density function max values and corresponding position, i.e. number of attractors.	110
8.2	Max number of attractors as a function of self-loops n_s	110
10.1	Summary of the experimental parameters.	133

Appendices

Appendix A

The impact of self-loops on BN attractor landscape

For the sake of clarity, in Section 8.4, we only presented the results concerning the CONST-OR configuration. Here, results for the other experimental configurations—namely AUGM-OR, CONST-RND and AUGM-RND—are reported.

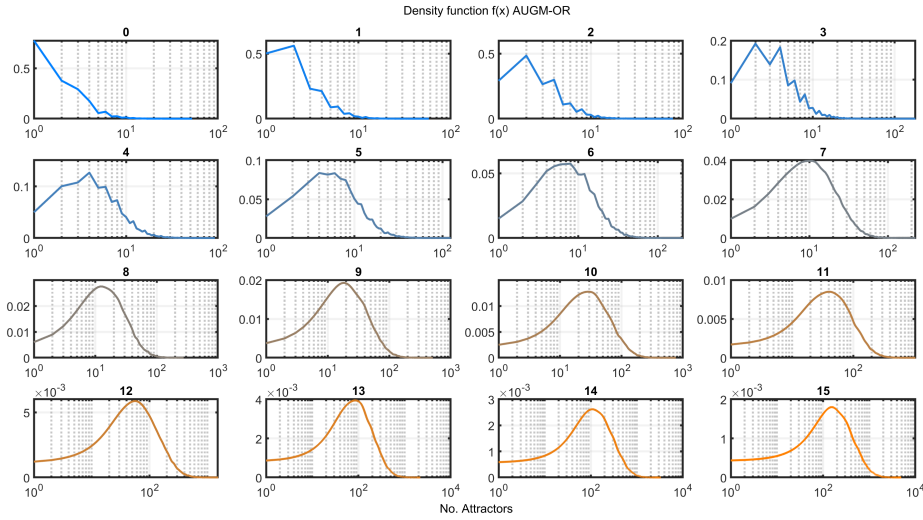


Figure A.1: Density function of the number of attractors for the AUGM-OR configuration. It has been computed from experimental data on $2 * 10^4$ runs with same n_s . The sixteen plots show how it changes with $n_s \in \{0, 1, \dots, 15\}$ (see plot title). Logarithmic scale for the x axis.

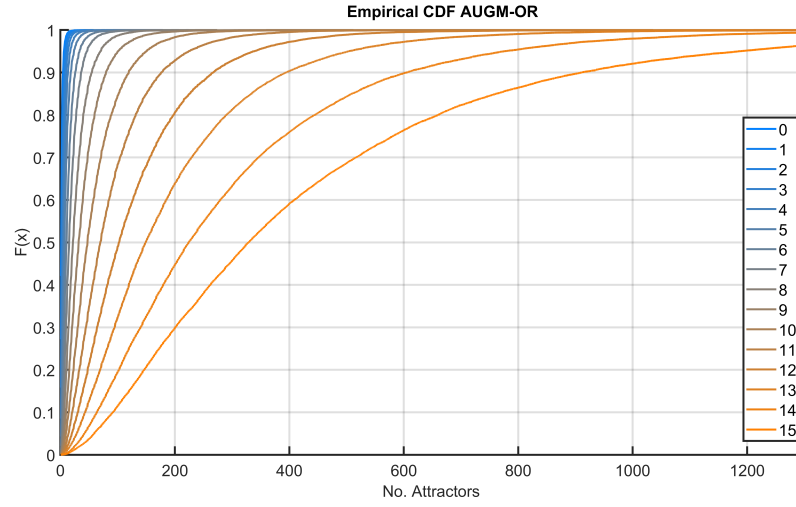


Figure A.2: Empirical cumulative distribution function (ECDF) of the number of attractors for the AUGM-OR configuration. It has been computed from experimental data on $2 \cdot 10^4$ runs with same n_s , with $n_s \in \{0, 1, \dots, 15\}$ (see legend). With respect to the CONST-OR configuration, we can observe that, if we compare curves with same n_s , 95% of networks are under the ECDF curve for a higher (almost doubled) number of attractors. For instance with $n_s = 15$ we have 95% of networks with $m \in [1, 1200]$ for AUGM-OR, against $m \in [1, 600]$ for CONST-OR.

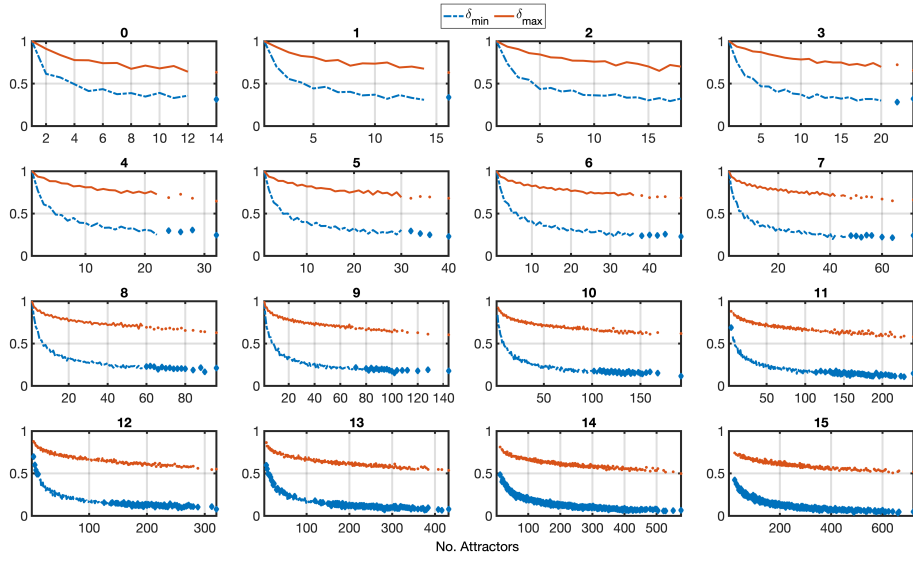


Figure A.3: δ_{min} and δ_{max} as a function of the number of attractors m for the AUGM-OR configuration. They have been computed from experimental data on $2 * 10^4$ runs with same n_s . The sixteen plots show how they change with $n_s \in \{0, 1, \dots, 15\}$ (see plot title). x axis limit changes with n_s from 1 to the largest number of attractors for which at least 30 sampled BNs have been found. Differently from the case of CONST-OR, we do not observe here a substantial increase of δ_{max} with gradually increasing number of self-loops in the network.

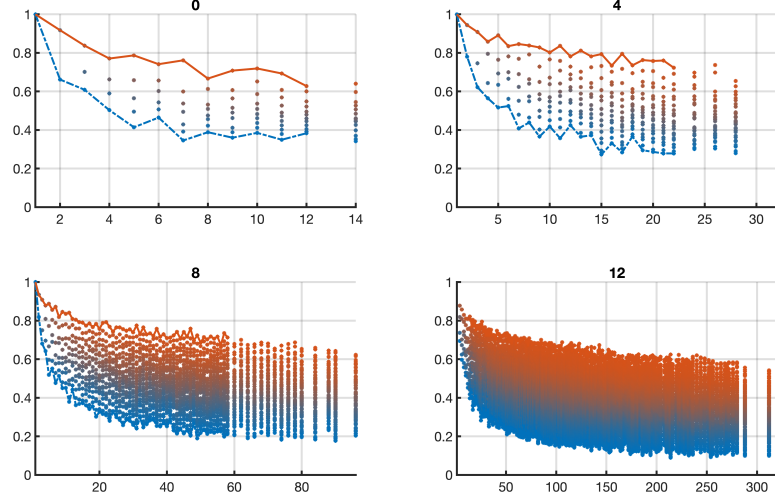


Figure A.4: δ_{min} , δ_{max} and ATM diagonal values as a function of the number of attractors for AUGM-OR configuration. Results are shown for $n_s \in \{0, 4, 8, 12\}$ (see title in each plot).

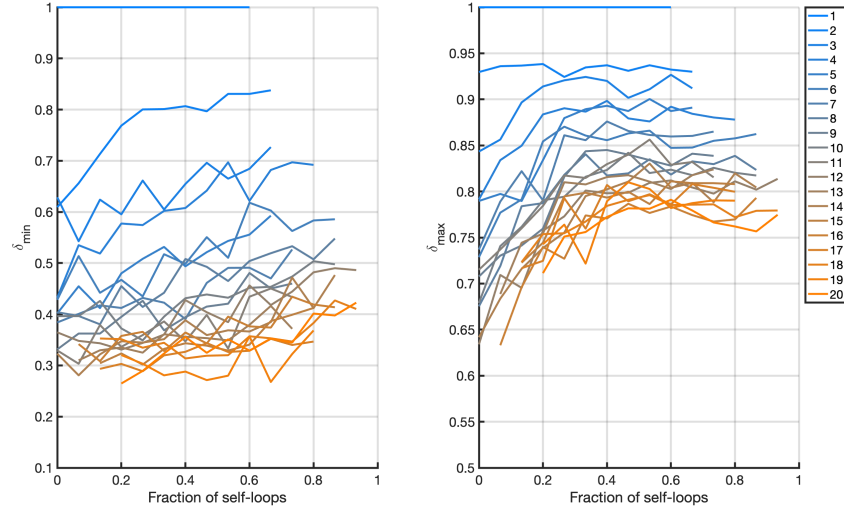


Figure A.5: δ_{min} (left) and δ_{max} (right) as a function of the fraction of nodes with a self-loop for AUGM-OR configuration. Each plot refers to the average values of δ_{min} (left) and δ_{max} (right) on 30 networks with the same number of attractors m , where $m \in \{1, \dots, 20\}$ (see legend).

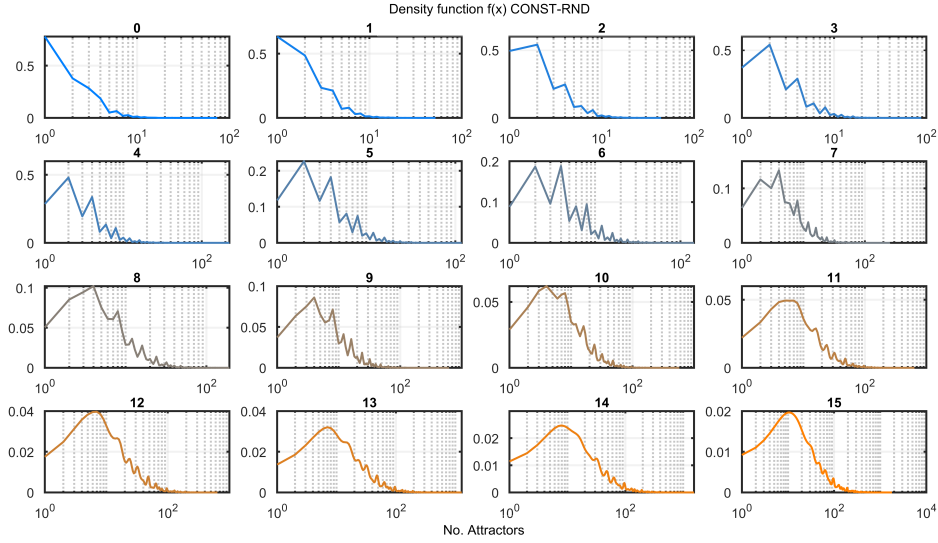


Figure A.6: Density function of the number of attractors for the CONST-RND configuration. It has been computed from experimental data on $2 * 10^4$ runs with same n_s . The sixteen plots show how it changes with $n_s \in \{0, 1, \dots, 15\}$ (see plot title). Logarithmic scale for the x axis.

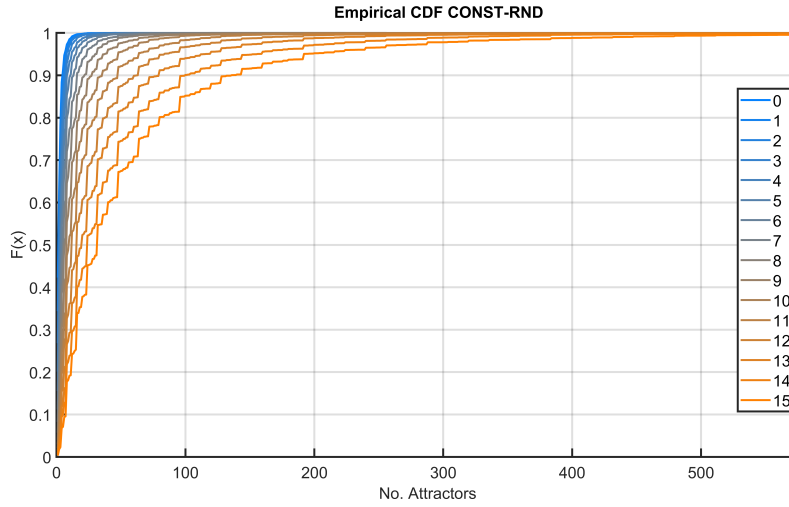


Figure A.7: Empirical cumulative distribution function (ECDF) of the number of attractors for the CONST-RND configuration. It has been computed from experimental data on $2 * 10^4$ runs with same n_s , with $n_s \in \{0, 1, \dots, 15\}$ (see legend). With respect to the CONST-OR configuration, we can observe that, if we compare curves with same n_s , 95% of networks are under the ECDF curve for a lower (about one third) number of attractors. For instance with $n_s = 15$ we have 95% of networks with $m \in [1, 200]$ for CONST-RND, against $m \in [1, 600]$ for CONST-OR.

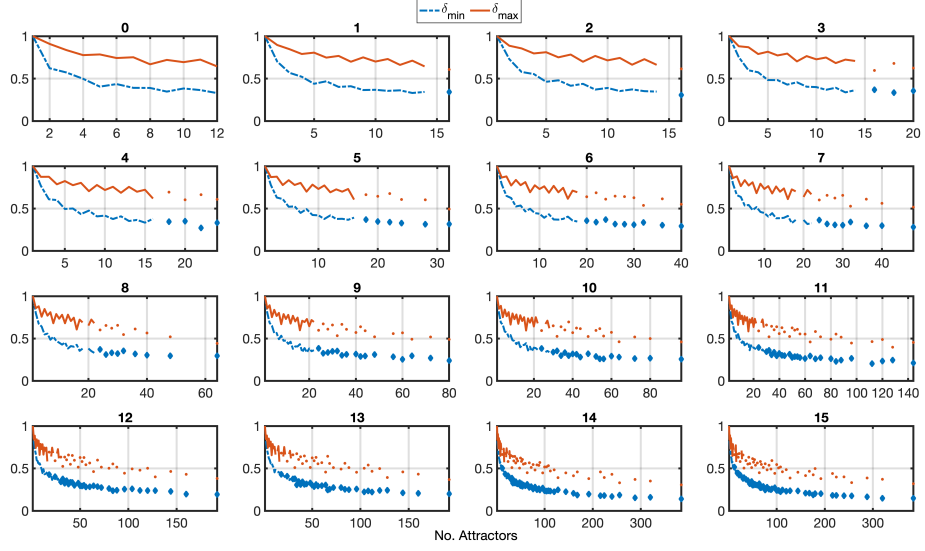


Figure A.8: δ_{min} and δ_{max} as a function of the number of attractors m for the CONST-RND configuration. They have been computed from experimental data on $2 * 10^4$ runs with same n_s . The sixteen plots show how they change with $n_s \in \{0, 1, \dots, 15\}$ (see plot title). x axis limit changes with n_s from 1 to the largest number of attractors for which at least 30 sampled BNs have been found. If compared across networks with the same number of attractors—but different n_s —maximum and minimum values of the ATM diagonal do not change, meaning that adding self-loops with the CONST-RND configuration does not provide neither an advantage nor a disadvantage in terms of attractor robustness.

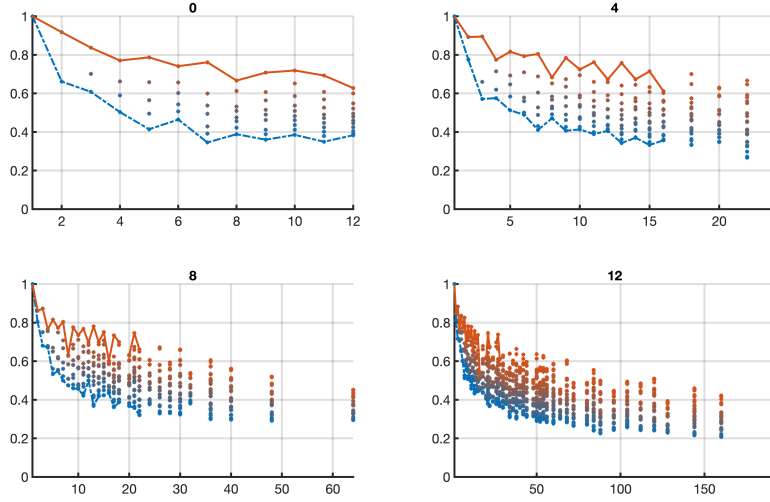


Figure A.9: δ_{min} , δ_{max} and ATM diagonal values as a function of the number of attractors for CONST-RND configuration. Results are shown for $n_s \in \{0, 4, 8, 12\}$ (see title in each plot).

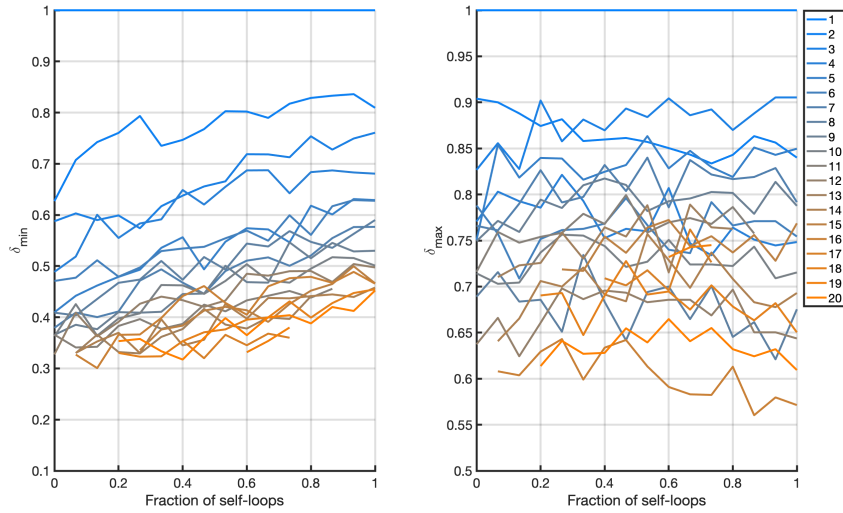


Figure A.10: δ_{min} (left) and δ_{max} (right) as a function of the fraction of nodes with a self-loop for CONST-RND configuration. Each plot refers to the average values of δ_{min} (left) and δ_{max} (right) on 30 networks with the same number of attractors m , where $m \in \{1, \dots, 20\}$ (see legend).

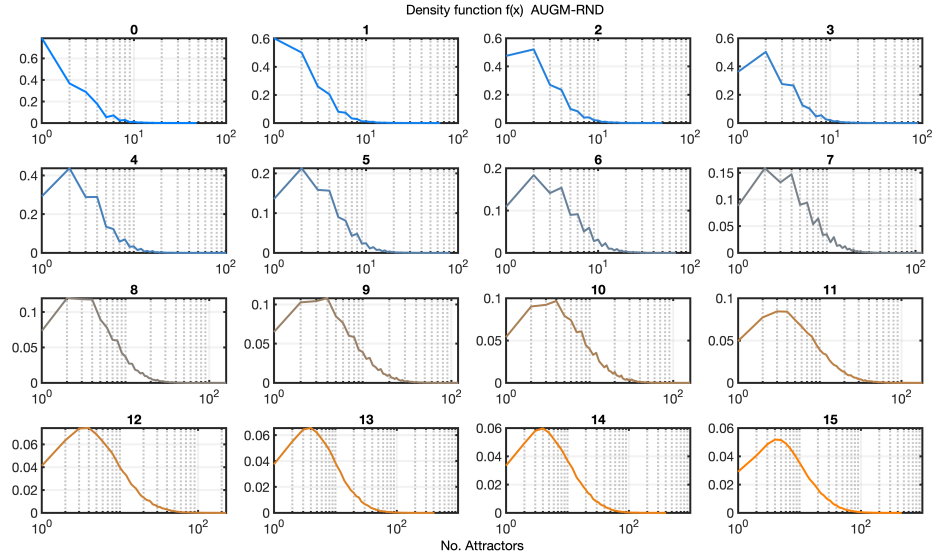


Figure A.11: Density function of the number of attractors for the AUGM-RND configuration. It has been computed from experimental data on $2 * 10^4$ runs with same n_s . The sixteen plots show how it changes with $n_s \in \{0, 1, \dots, 15\}$ (see plot title).

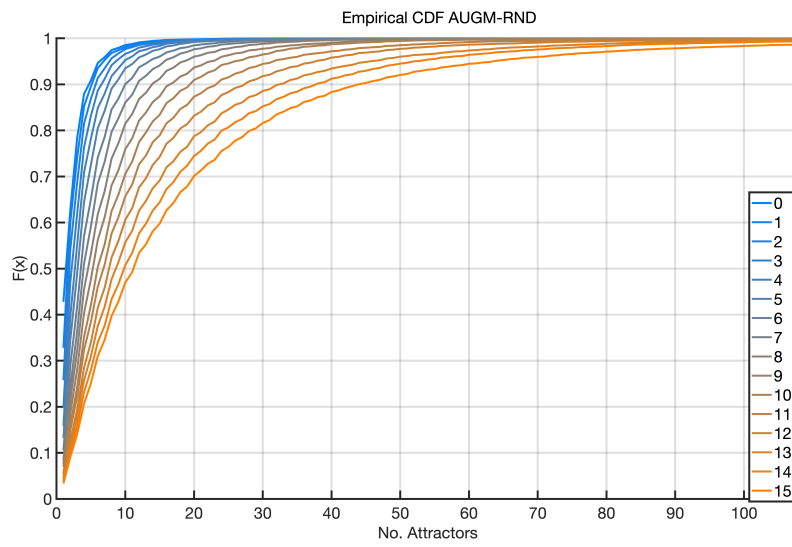


Figure A.12: Empirical cumulative distribution function (ECDF) of the number of attractors for the AUGM-RND configuration. It has been computed from experimental data on $2 \cdot 10^4$ runs with same n_s , with $n_s \in \{0, 1, \dots, 15\}$ (see legend). Even though adding self-loops results in an increased number of attractors also in the case of AUGM-RND configuration, changes in the distribution of the number of attractors are less substantive with respect to the previous configurations.

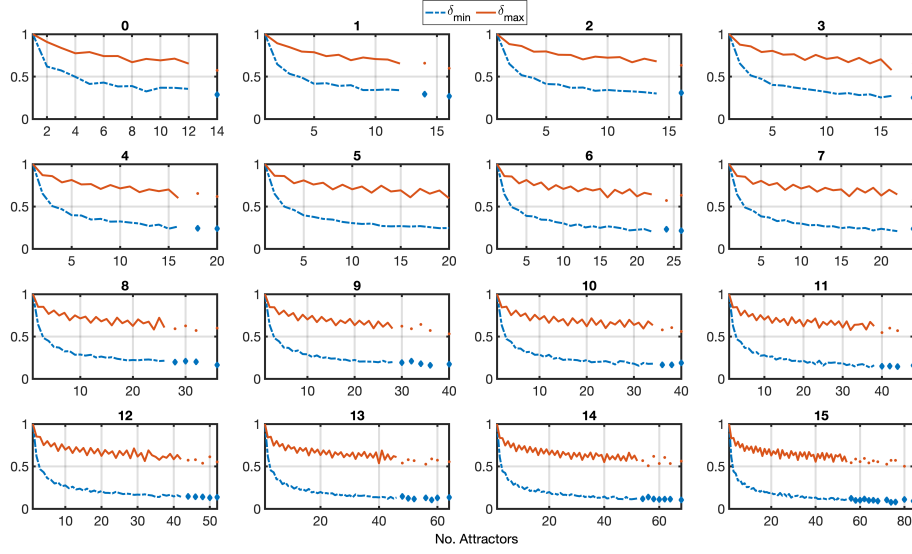


Figure A.13: δ_{min} and δ_{max} as a function of the number of attractors m for the AUGM-RND configuration. They have been computed from experimental data on $2 * 10^4$ runs with same n_s . The sixteen plots show how they change with $n_s \in \{0, 1, \dots, 15\}$ (see plot title). x axis limit changes with n_s from 1 to the largest number of attractors for which at least 30 sampled BNs have been found. If compared across networks with the same number of attractors—but different n_s —maximum and minimum values of the ATM diagonal do not change, meaning that adding self-loops with the AUGM-RND configuration does not provide neither an advantage nor a disadvantage in terms of attractor robustness.

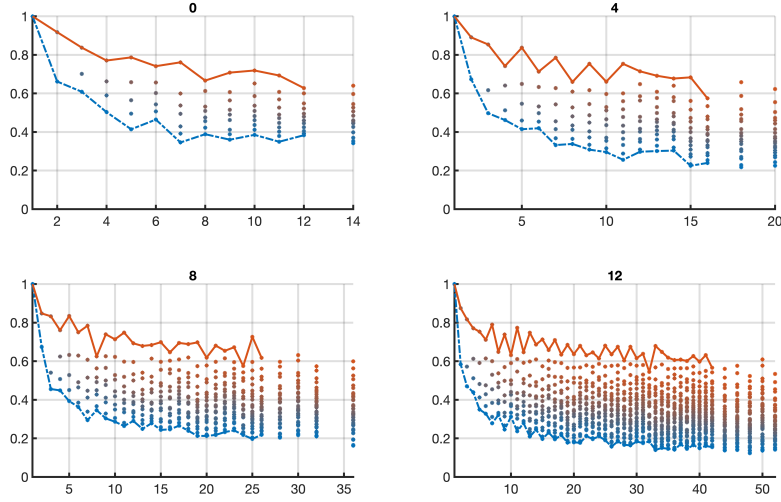


Figure A.14: δ_{min} , δ_{max} and ATM diagonal values as a function of the number of attractors for AUGM-RND configuration. Results are shown for $n_s \in \{0, 4, 8, 12\}$ (see title in each plot).

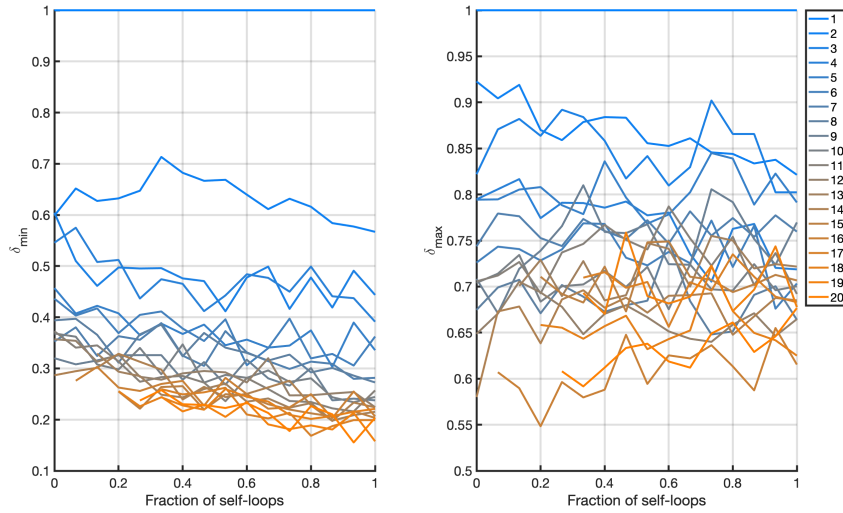


Figure A.15: δ_{min} (left) and δ_{max} (right) as a function of the fraction of nodes with a self-loop for AUGM-RND configuration. Each plot refers to the average values of δ_{min} (left) and δ_{max} (right) on 30 networks with the same number of attractors m , where $m \in \{1, \dots, 20\}$ (see legend).

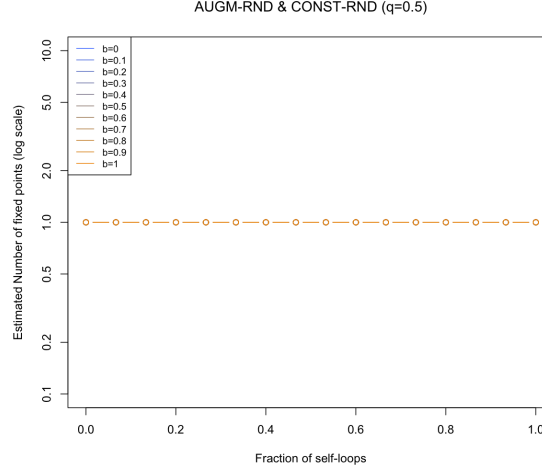


Figure A.16: Analytical model results for AUGM-RND and CONST-RND cases with varying bias as a function of the fraction of self-loops. Since the average outcome of the Boolean functions in RBN ensembles of random functions follows the bias parameter, the analytical model for the estimation of the average number of fixed points for the RND cases can be reduced to the following formula $[bq + (1 - b)(1 - q)]^n$ (b = bias, n = nodes number, q = probability of assigning the value 1 to a node).

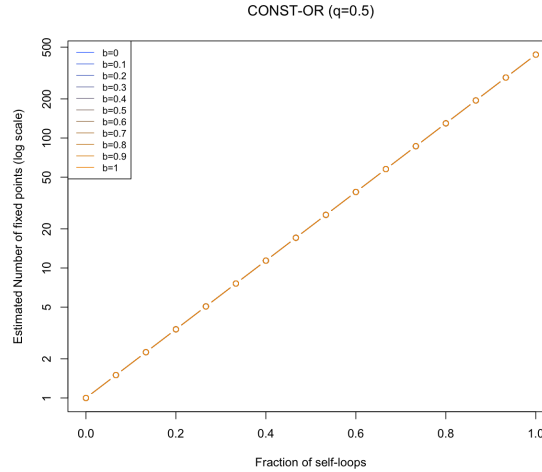


Figure A.17: Analytical model results for the CONST-OR case with varying bias as a function of the fraction of self-loops.

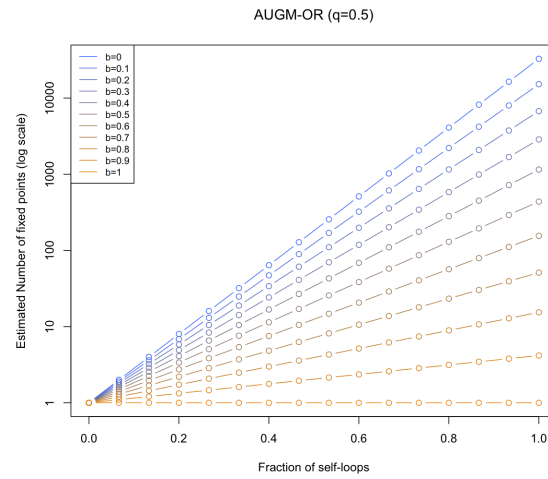


Figure A.18: Analytical model results for the cases AUGM-OR with varying bias as a function of the fraction of self-loops.